

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 13:43:01 ; Search time 3357 Seconds
(without alignments)
6029.549 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctatccacct.....taataaacagttaaaagctg 467

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

GenEmbl.*

1: gb_bat.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_man.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|------------|--------------------|
| 1 | 467 | 100.0 | 494 | 6 | AX092298 | Sequence |
| 2 | 467 | 100.0 | 494 | 6 | AX696953 | Sequence |
| 3 | 467 | 100.0 | 495 | 9 | AY358657 | Homo sapi |
| 4 | 467 | 100.0 | 2673 | 9 | BC021104 | Sequence |
| 5 | 467 | 100.0 | 50000 | 6 | AR211704 | Sequence |
| 6 | 467 | 100.0 | 151152 | 9 | HS454M7 | Human DNA |
| 7 | 467 | 100.0 | 206618 | 9 | AF195953 | Homo sapi |
| 8 | 454 | 97.2 | 2243 | 6 | BD156791 | Primer fo |
| 9 | 454 | 97.2 | 2243 | 9 | AK001855 | Homo sapi |
| 10 | 454 | 97.2 | 2243 | 9 | AK001855 | Homo sapi |
| 11 | 419.2 | 89.8 | 458 | 6 | AX332625 | Sequence |
| 12 | 419.2 | 89.8 | 458 | 6 | AX332625 | Sequence |
| 13 | 391.2 | 83.8 | 403 | 6 | AX874329 | Sequence |
| 14 | 391.2 | 83.8 | 403 | 6 | BD154391 | Primer fo |
| 15 | 151.8 | 32.5 | 171310 | 2 | AC127934 | Rattus no |
| 16 | 151.8 | 32.5 | 208810 | 2 | AC136814 | Rattus no |
| 17 | 129.8 | 27.8 | 3143 | 10 | BC020015 | Mus muscu |
| 18 | 129.8 | 27.8 | 118604 | 10 | AL714010 | Mouse DNA |
| 19 | 129.8 | 27.8 | 189982 | 10 | AL672274 | Mouse DNA |
| 20 | 129.8 | 27.8 | 203805 | 10 | AC055817 | Mus muscu |
| 21 | 127 | 27.2 | 598 | 9 | H0MZB38F03 | Homo sapi |
| 22 | 74.2 | 15.9 | 255 | 6 | AX211242 | Sequence |
| 23 | 46.4 | 9.9 | 73778 | 6 | AX344562 | Sequence |
| 24 | 46.4 | 9.9 | 349980 | 6 | AX344561 | Sequence |
| 25 | 45 | 9.6 | 134971 | 2 | AC116367 | Oryza sat |
| 26 | 45 | 9.6 | 212887 | 2 | AC134496 | Rattus no |
| 27 | 45 | 9.6 | 272030 | 2 | AC105482 | Rattus no |
| 28 | 44.2 | 9.5 | 71781 | 2 | AC101159 | Mus muscu |
| 29 | 44.2 | 9.5 | 213417 | 2 | AC119515 | Rattus no |
| 30 | 44.2 | 9.5 | 242655 | 2 | AC107465 | Rattus no |
| 31 | 44.2 | 9.5 | 243369 | 10 | AL606973 | Mouse DNA |
| 32 | 44 | 9.4 | 173254 | 2 | AC146436 | Pan trogl |
| 33 | 43.8 | 9.4 | 903 | 3 | AF441245 | Nephila c |
| 34 | 43.8 | 9.4 | 155665 | 2 | AC133581 | Mus muscu |
| 35 | 43.2 | 9.3 | 119790 | 9 | HSDJ989D7 | Human DNA |
| 36 | 43 | 9.2 | 59414 | 9 | AC093112 | Homo sapi |
| 37 | 43 | 9.2 | 343188 | 2 | AC114853 | Rattus no |
| 38 | 42.8 | 9.2 | 224334 | 2 | AC096041 | Rattus no |
| 39 | 42.6 | 9.1 | 575 | 10 | MUSTCBVC1 | Mouse germ |
| 40 | 42.6 | 9.1 | 982 | 10 | MUSTCRBBA | Mouse beta- |
| 41 | 42.6 | 9.1 | 2708 | 3 | DROZSTA | Drosophila |
| 42 | 42.6 | 9.1 | 145052 | 10 | AL603889 | Mouse DNA |
| 43 | 42.6 | 9.1 | 199221 | 2 | AC128156 | Rattus no |
| 44 | 42.6 | 9.1 | 250611 | 10 | MMAE000663 | Mus muscu |
| 45 | 42.6 | 9.1 | 261600 | 10 | AC125228 | Mus muscu |
| 46 | 42.4 | 9.1 | 139786 | 10 | AC124343 | Mus muscu |
| 47 | 42.4 | 9.1 | 155276 | 2 | AC091992 | Homo sapi |
| 48 | 42.4 | 9.1 | 202733 | 10 | AC122805 | Mus muscu |
| 49 | 42.4 | 9.1 | 203124 | 2 | AC131668 | Mus muscu |
| 50 | 42.4 | 9.1 | 234228 | 2 | AC125852 | Rattus no |
| 51 | 42.2 | 9.0 | 50783 | 2 | AC120601 | Rattus no |
| 52 | 42.2 | 9.0 | 192814 | 9 | AC007597 | Homo sapi |
| 53 | 42.2 | 9.0 | 214786 | 9 | AC007610 | Homo sapi |
| 54 | 42.2 | 9.0 | 227164 | 2 | AC103418 | Rattus no |
| 55 | 42.2 | 9.0 | 227663 | 2 | AC097301 | Rattus no |
| 56 | 42.2 | 9.0 | 240105 | 2 | AC117339 | Rattus no |
| 57 | 42.2 | 9.0 | 241730 | 2 | AC103283 | Rattus no |
| 58 | 42.2 | 9.0 | 283782 | 2 | AC096338 | Rattus no |
| 59 | 42 | 9.0 | 55116 | 9 | AP005854 | Homo sapi |
| 60 | 41.8 | 8.9 | 217240 | 5 | BX004795 | Zebrafish |
| 61 | 41.6 | 8.9 | 110000 | 2 | AC105872 | Continuation (3 of |
| 62 | 41.6 | 8.9 | 142278 | 9 | AC005177 | Continuation (3 of |
| 63 | 41.6 | 8.9 | 241865 | 2 | AC126834 | Rattus no |
| 64 | 41.6 | 8.9 | 315654 | 2 | AC118159 | Rattus no |
| 65 | 41.4 | 8.9 | 1585 | 5 | AY225852 | Ctenophar |

Pred. No. is the number of results predicted by chance to have a


```

QY 181 TGTCTCTGCTCATGGCTGCGCATGACCTGAGCCAGCCAGCCCGCTGGGGAAGGG 240
Db 181 TGTCTCTGCTCATGGCTGCGCATGACCTGAGCCAGCCAGCCCGCTGGGGAAGGG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
QY 301 GGCTAGGGGGGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATACAA 360
Db 301 GGCTAGGGGGGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATACAA 360
QY 361 AGATATTAAAGGCGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATACAA 420
Db 361 AGATATTAAAGGCGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATACAA 420
QY 421 AGATTGTTCTGTGTAATATGTTCTTTAATAAACAAGTAAAGCTG 467
Db 421 AGATTGTTCTGTGTAATATGTTCTTTAATAAACAAGTAAAGCTG 467

RESULT 3
AY358657 495 bp mRNA linear PRI 03-OCT-2003
LOCUS Homo sapiens clone DNA56962 APELIN (UNQ471) mRNA, complete cds.
ACCESSION AY358657
VERSION AY358657.1 GI:37182435
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J.,
Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B.,
Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E.,
Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S.,
Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C.,
Schoenfeld,J., Seehagiri,S., Simmons,L., Singh,J., Smith,V.,
Stinson,J., Vagstad,A., Vandlen,R., Watanabe,C., Wleand,D., Woods,K.,
Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z.,
Goddard,A., Wood,W.I. and Godowski,P.
The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
Effort to Identify Novel Human Secreted and Transmembrane Proteins:
A Bioinformatics Assessment
Genome Res. 13 (10), 2265-2270 (2003)
JOURNAL 12975309
PUBMED 2 (bases 1 to 495)
AUTHORS Clark,H.F.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Department of Bioinformatics, Genentech,
Inc., 1 DNA Way, South San Francisco, CA 94080, USA
FEATURES
source
1. 495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_type="taxon:9606"
/clone="DNA56962"
1. 495
/locus_tag="UNQ471"
40. 261
/locus_tag="UNQ471"
/feature="PR0831"
/codon_start=1
/product="APELIN"
/protein_id="AAQ89020.1"
/db_xref="GI:37182436"
/translators="MILLIUMILLIKSGLENGIAGQKVSSATDAPIRDNAFPFP
SFCLLPHPAMTCSQAQPRGEKVGDG"

Query Match 100.0%; Score 467; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.3e-113;

```

```

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTGGCTATCCACCTCCCGCCAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
Db 1 CAATGTTGGCTATCCACCTCCCGCCAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
QY 61 CTGCTGCTGCTGCTGTTAAAGGCTCATGCTGAGTGGGAGCTGGTGGTGGCCAGAAA 120
Db 61 CTGCTGCTGCTGCTGTTAAAGGCTCATGCTGAGTGGGAGCTGGTGGTGGCCAGAAA 120
QY 121 GTCTCTTCTGCCACTGACGCCCGCCCATCAGGAGTGGGCGCTTCTTCCCCCTTCTTCTG 180
Db 121 GTCTCTTCTGCCACTGACGCCCGCCCATCAGGAGTGGGCGCTTCTTCCCCCTTCTTCTG 180
QY 181 TGTCTCTGCTCATGGCTGCGCATGACCTGAGCCAGCCAGCCCGCTGGGGAAGGG 240
Db 181 TGTCTCTGCTCATGGCTGCGCATGACCTGAGCCAGCCAGCCCGCTGGGGAAGGG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
QY 301 GGCTAGGGGGGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATACAA 360
Db 301 GGCTAGGGGGGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATACAA 360
QY 361 AGATATTAAAGGCGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATGTA 420
Db 361 AGATATTAAAGGCGCTGCTTAATTTAAAGTGGTGTGTTATGATTCCTACTAATTATGTA 420
QY 421 AGATTGTTCTGTGTAATATGTTCTTTAATAAACAAGTAAAGCTG 467
Db 421 AGATTGTTCTGTGTAATATGTTCTTTAATAAACAAGTAAAGCTG 467

RESULT 4
BC021104 2673 bp mRNA linear PRI 07-OCT-2003
LOCUS Homo sapiens apelin, mRNA (cDNA clone MGC:31846 IMAGE:4586949),
complete cds.
ACCESSION BC021104
VERSION BC021104.1 GI:18088893
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2673)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,D., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Casavant,T.L.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalius,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL 22388257
MEDLINE 12477932
PUBMED 2 (bases 1 to 2673)
REFERENCE Strausberg,R.
AUTHORS

```

TITLE
JOURNAL

REMARK
COMMENT

Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIR-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
BC Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiac, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 40 Row: f Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21314667.
Location/Qualifiers
1. .2673
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:31846 IMAGE:4586949"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH MGC 14"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. .2673
/gene="APLN"
/db_xref="LocusID:8862"
/db_xref="MIM:300297"
420. .788
/codon_start=1
/product="apelin"
/protein_id="AAH21104.1"
/db_xref="GI:18088894"
/db_xref="LocusID:8862"
/translations="MSATWCSPEGQMGQPGREVGNAGSPASPIRDPCLSEAGL
KGPPSAHPRRLCLHLVCFSGSLTSIQLSPRTCCSHQWALFSPACFFQWRAPGCSL
DSSRSRLTRPVHLPSPSLD"

ORIGIN

Query Match 100.0%; Score 467; DB 9; Length 2673;
Best Local Similarity 100.0%; Pred. No. 5.7e-113;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
Db 2188 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 2247

QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGCTGCTGCTGCTGCTGCTG 120
Db 2248 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGCTGCTGCTGCTGCTGCTG 2307

QY 121 GTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Db 2308 GTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2367

QY 181 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 2368 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2427

QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAAAGAGGCTAGTGGGTG 300
Db 2428 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAAAGAGGCTAGTGGGTG 2487

QY 301 GGCTAGGGGGCTGCTTATTTAAAGTGGTGGTATGATCTTATATCTAATTTATACAA 360
Db 2488 GGCTAGGGGGCTGCTTATTTAAAGTGGTGGTATGATCTTATATCTAATTTATACAA 2547

QY 361 AGATATTAAAGCCCTGTTTCAATTAAGAAATTTCCCTTCCCTGCTGCTCAATGTTGTAA 420
Db 2548 AGATATTAAAGCCCTGTTTCAATTAAGAAATTTCCCTTCCCTGCTGCTCAATGTTGTAA 2607

QY 421 AGATGTTTGTGTAAATATGCTTTTATAATAACAGTTAAAGCTG 467
Db 2608 AGATGTTTGTGTAAATATGCTTTTATAATAACAGTTAAAGCTG 2654

RESULT 5
AR211704/c
LOCUS
DEFINITION Sequence 4 from patent US 6399349. DNA linear PAT 20-JUN-2002
ACCESSION AR211704
VERSION AR211704.1 GI:21515097
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 50000)
AUTHORS Ryan, J.W. and Sprinkle, T. Joe. Curtis.
TITLE Human aminopeptidase P gene.
JOURNAL Patent: US 6399349-A 4 04-JUN-2002;
FEATURES
source
1. .50000
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 467; DB 6; Length 50000;
Best Local Similarity 100.0%; Pred. No. 6.4e-113;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
Db 1518 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 1459

QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGCTGCTGCTGCCAGAAA 120
Db 1458 CTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGCTGCTGCTGCCAGAAA 1399

QY 121 GTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Db 1398 GTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1339

QY 181 TGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 1338 TGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279

QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGCTAGTGGGTG 300
Db 1278 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGCTAGTGGGTG 1219

QY 301 GGCTAGGGGGCTGCTTATTTAAAGTGGTGGTATGATTTTATATCTTATATCTAATTTATACAA 360
Db 1218 GGCTAGGGGGCTGCTTATTTAAAGTGGTGGTATGATTTTATATCTTATATCTAATTTATACAA 1159

QY 361 AGATATTAAAGCCCTGTTTCAATTAAGAAATTTCCCTTCCCTGCTGCTCAATGTTGTAA 420
Db 1158 AGATATTAAAGCCCTGTTTCAATTAAGAAATTTCCCTTCCCTGCTGCTCAATGTTGTAA 1099

QY 421 AGATGTTTGTGTAAATATGCTTTTATAATAACAGTTAAAGCTG 467
Db 1098 AGATGTTTGTGTAAATATGCTTTTATAATAACAGTTAAAGCTG 1052

RESULT 6
 HS454M7/c
 LOCUS
 DEFINITION
 Human DNA sequence from clone RP3-454M7 on chromosome Xq25-26.3, complete sequence.
 ACCESSION
 AL022162
 VERSION
 AL022162.1
 GI:3171881
 KEYWORDS
 HTG.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 151152)
 AUTHORS
 Pavitt R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT
 On Jun 2, 1998 this sequence version replaced gi:2965945.

 Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SwissProt; Tr; TrEMBL; Wp; WormPeP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>
 RP3-454M7 is from the library RP3-454M7-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP3-454M7.
 FEATURES
 Location/Qualifiers
 1..151152
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="RZPD:RP3-454M7" /db_xref="taxon:9606"
 /chromosome="X"
 /map="q25-26.3"
 /clone="RP3-454M7"
 /clone_lib="RP3-454M7-3"
 /gene="OCRL1"
 767..35998
 join(<767..877,1303..1392,2075..2195,2282..2443,3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,10696..10805,12706..12841,18582..18692,19339..19504,19759..19994,30444..30560,31621..31705,32328..32455,33287..33398,33588..33712)
 /gene="OCRL1"
 /product="du454M7.1.1 (Low Oculocerebrorenal Syndrome)"
 /protein_id="CAA18150.1"
 /db_xref="GI:4160528"
 /translation="GCKIRVQGDWIRERRFEIPDEHCLKFLSAVLAQAQKASQLLVP
 ECKDSSWYQKLDTKDQKPSVSGLLGFDFNFSMNLDKKINSQNTGTHREPPPPF
 SYNKMLPREKEASNKQKPVNTMRKLFVNTQSGRGLIKHILAKREKVEVNTQTF
 RFVGTNVNGSQSGGLEPWLNCNPPDIYICIGFQELDLSTEAFPFVSVEKQES
 MAVERGLHSKAKYKQVLRVGLMMLIFARKQCYRIADIATETVGTGIMGKNK
 GAVRFVPHNTFCIVNSHLAAHVDFERRNQDYKICARMSFVVPNTLPQNLIMKH
 EVVILWGLDNLVRLCMPDANEKSLINKDLQKLFQDLNLTOROKKAFVDFNEGEIK
 FIPTYKDSKTRDWDSSGKCRVPACDRILWGTNTVNLNRYSHMELTSDHKPSYAL
 FHTYKVDYRKYKVFSDSVIRMDRMENDFLPSLESRFVFNVEKFRQIKQKEQ
 ISNNGQVCFHSFIPKNDQSQCCKPWLRAPEFGLPEFNETVDSIDVYVSKDSVIL
 NSGDKIDILVHLDRGKDYFLTISGNLYPFCFSGTSLALCMKRPRIREVPTGLM
 LEEDSFLKESLQWPLDEGASERPLQVKEIWLVDHFKYACHQEDLFTQGMQ
 BELQIIDLCLDTSIPETIPGSHNSVAEALLIFLEALPEPVI CYELYQCLDSAYDPRI
 CQVTSOLPRCHNVRYLMAFLRELLKFSYNSVANNMIAITFTSLLRPPNPPNLMAR
 QTESDQRAIQELGLLGSEED"
 join(<769..877,1303..1392,2075..2195,2282..2443,3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,10696..10805,12706..12841,18582..18692,19339..19504,19759..19994,30444..30560,31621..31705,32328..32455,33287..33398,33588..33712)
 /gene="OCRL1"
 /note="variant 1"
 /codon_start=1
 /evidence=not experimental
 /product="du454M7.1.1 (Low Oculocerebrorenal Syndrome)"
 /protein_id="CAA18151.1"
 /db_xref="GI:4160528"
 /translation="GCKIRVQGDWIRERRFEIPDEHCLKFLSAVLAQAQKASQLLVP
 ECKDSSWYQKLDTKDQKPSVSGLLGFDFNFSMNLDKKINSQNTGTHREPPPPF
 SYNKMLPREKEASNKQKPVNTMRKLFVNTQSGRGLIKHILAKREKVEVNTQTF
 RFVGTNVNGSQSGGLEPWLNCNPPDIYICIGFQELDLSTEAFPFVSVEKQES
 MAVERGLHSKAKYKQVLRVGLMMLIFARKQCYRIADIATETVGTGIMGKNK
 GAVRFVPHNTFCIVNSHLAAHVDFERRNQDYKICARMSFVVPNTLPQNLIMKH
 EVVILWGLDNLVRLCMPDANEKSLINKDLQKLFQDLNLTOROKKAFVDFNEGEIK
 FIPTYKDSKTRDWDSSGKCRVPACDRILWGTNTVNLNRYSHMELTSDHKPSYAL
 FHTYKVDYRKYKVFSDSVIRMDRMENDFLPSLESRFVFNVEKFRQIKQKEQ
 ISNNGQVCFHSFIPKNDQSQCCKPWLRAPEFGLPEFNETVDSIDVYVSKDSVIL
 NSGDKIDILVHLDRGKDYFLTISGNLYPFCFSGTSLALCMKRPRIREVPTGLM
 LEEDSFLKESLQWPLDEGASERPLQVKEIWLVDHFKYACHQEDLFTQGMQ
 BELQIIDLCLDTSIPETIPGSHNSVAEALLIFLEALPEPVI CYELYQCLDSAYDPRI
 CQVTSOLPRCHNVRYLMAFLRELLKFSYNSVANNMIAITFTSLLRPPNPPNLMAR
 QTESDQRAIQELGLLGSEED"
 join(<769..877,1303..1392,2075..2195,2282..2443,3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,10696..10805,12706..12841,18582..18692,19339..19504,19759..19994,30444..30560,31621..31705,32328..32455,33287..33398,33588..33712)
 /gene="OCRL1"
 /note="variant 1"
 /codon_start=1
 /evidence=not experimental
 /product="du454M7.1.1 (Low Oculocerebrorenal Syndrome)"
 /protein_id="CAA18151.1"
 /db_xref="GI:4160528"
 /translation="GCKIRVQGDWIRERRFEIPDEHCLKFLSAVLAQAQKASQLLVP
 ECKDSSWYQKLDTKDQKPSVSGLLGFDFNFSMNLDKKINSQNTGTHREPPPPF
 SYNKMLPREKEASNKQKPVNTMRKLFVNTQSGRGLIKHILAKREKVEVNTQTF
 RFVGTNVNGSQSGGLEPWLNCNPPDIYICIGFQELDLSTEAFPFVSVEKQES
 MAVERGLHSKAKYKQVLRVGLMMLIFARKQCYRIADIATETVGTGIMGKNK
 GAVRFVPHNTFCIVNSHLAAHVDFERRNQDYKICARMSFVVPNTLPQNLIMKH
 EVVILWGLDNLVRLCMPDANEKSLINKDLQKLFQDLNLTOROKKAFVDFNEGEIK
 FIPTYKDSKTRDWDSSGKCRVPACDRILWGTNTVNLNRYSHMELTSDHKPSYAL
 FHTYKVDYRKYKVFSDSVIRMDRMENDFLPSLESRFVFNVEKFRQIKQKEQ
 ISNNGQVCFHSFIPKNDQSQCCKPWLRAPEFGLPEFNETVDSIDVYVSKDSVIL
 NSGDKIDILVHLDRGKDYFLTISGNLYPFCFSGTSLALCMKRPRIREVPTGLM
 LEEDSFLKESLQWPLDEGASERPLQVKEIWLVDHFKYACHQEDLFTQGMQ
 BELQIIDLCLDTSIPETIPGSHNSVAEALLIFLEALPEPVI CYELYQCLDSAYDPRI
 CQVTSOLPRCHNVRYLMAFLRELLKFSYNSVANNMIAITFTSLLRPPNPPNLMAR
 QTESDQRAIQELGLLGSEED"
 join(<769..877,1303..1392,2075..2195,2282..2443,3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,10696..10805,12706..12841,18582..18692,19339..19504,19759..19994,30444..30560,31621..31705,32328..32455,33287..33398,33588..33712)
 /gene="OCRL1"
 /note="variant 1"
 /codon_start=1
 /evidence=not experimental
 /product="du454M7.1.1 (Low Oculocerebrorenal Syndrome)"
 /protein_id="CAA18150.1"
 /db_xref="GI:3171882"
 /translation="GCKIRVQGDWIRERRFEIPDEHCLKFLSAVLAQAQKASQLLVP
 ECKDSSWYQKLDTKDQKPSVSGLLGFDFNFSMNLDKKINSQNTGTHREPPPPF
 SYNKMLPREKEASNKQKPVNTMRKLFVNTQSGRGLIKHILAKREKVEVNTQTF
 RFVGTNVNGSQSGGLEPWLNCNPPDIYICIGFQELDLSTEAFPFVSVEKQES
 MAVERGLHSKAKYKQVLRVGLMMLIFARKQCYRIADIATETVGTGIMGKNK
 GAVRFVPHNTFCIVNSHLAAHVDFERRNQDYKICARMSFVVPNTLPQNLIMKH
 EVVILWGLDNLVRLCMPDANEKSLINKDLQKLFQDLNLTOROKKAFVDFNEGEIK


```

RESULT 9
BD156791
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
OS Homo sapiens (human)
PD JP 2002191363-A/11634
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,ETSUO NISHIKAWA,KOJI HAYASHI,KAORU SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,ETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH key
FT CDS
Location/Qualifiers
1..2243
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
ORIGIN
Query Match 97.2%; Score 454; DB 6; Length 2243;
Best Local Similarity 99.4%; Pred. No. 1.6e-109;
Matches 467; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 CAATGTTTGGCTATCCACCTCCCAAGGCCCTTTTACCTATGCTGCTGCTAAAG---CTG 57
DB 1774 CAATGTTTGGCTATCCACCTCCCAAGGCCCTTTTACCTATGCTGCTGCTAAAGCTACTG 1833

QY 58 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 117
DB 1834 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 1893

QY 118 AAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCGCTTCTTTCCCGCTTCCTTT 177
DB 1894 AAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCGCTTCTTTCCCGCTTCCTTT 1953

QY 178 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 237
DB 1954 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 2013

QY 238 GGGGAGAAAGTGGGGATGGCTAAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTGG 297
DB 2014 GGGGAGAAAGTGGGGATGGCTTAAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTGG 2073

QY 298 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTATTATGATTTCTTAACTAAATTATA 357
DB 2074 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTATTATGATTTCTTAACTAAATTATA 2133

QY 358 CAAAGATATTAAGCCCTGTTTAAAGAAATGTTCCCTTCCCTCGTGTCAATGTTTG 417
DB 2134 CAAAGATATTAAGCCCTGTTTAAAGAAATGTTCCCTTCCCTCGTGTCAATGTTTG 2193

QY 418 TAAAGATTGTTCTGTGTAATATGTTCTTTTATAATAACAGTTAAAGCTG 467

```

```

DB 2194 TAAAGATTGTTCTGTGTAATATGTTCTTTTATAATAACAGTTAAAGCTG 2243

RESULT 10
AK001855
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
Query Match 97.2%; Score 454; DB 9; Length 2243;
Best Local Similarity 99.4%; Pred. No. 1.6e-109;
Matches 467; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 CAATGTTTGGCTATCCACCTCCCAAGGCCCTTTTACCTATGCTGCTGCTAAAG---CTG 57
DB 1774 CAATGTTTGGCTATCCACCTCCCAAGGCCCTTTTACCTATGCTGCTGCTAAAGCTACTG 1833

QY 58 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 117
DB 1834 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 1893

QY 118 AAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCGCTTCTTTCCCGCTTCCTTT 177
DB 1894 AAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCGCTTCTTTCCCGCTTCCTTT 1953

QY 178 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 237

```


Db 1954 CTGTGTCCTGCTCATCGCCCTGCCATGACCTGGAGCCAGCCCGCCGCGGAA 2013
Qy 238 GGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGAAACAGAGAGGTAGTGG 297
Db 2014 GGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGAAACAGAGAGGTAGTGG 2073
Qy 298 GTGGGCTAGGGGGCTGCTTATTAAGTGGTGTGTTTATGATCTTATTAATTTATA 357
Db 2074 GTGGGCTAGGGGGCTGCTTATTAAGTGGTGTGTTTATGATCTTATTAATTTATA 2133
Qy 358 CAAAGATATTAAAGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTGTTCAATGTTTG 417
Db 2134 CAAAGATATTAAAGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTGTTCAATGTTTG 2193
Qy 418 TAAAGATTGTTCTGTGTAATAATGTTCTTTATAATAAACAGTTAAAGCTG 467
Db 2194 TAAAGATTGTTCTGTGTAATAATGTTCTTTATAATAAACAGTTAAAGCTG 2243

RESULT 11
AX332625/c
LOCUS AX332625 458 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3134 from Patent WO0194629.
ACCESSION AX332625
VERSION AX332625.1 GI:18123259
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 3134 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source Location/Qualifiers
1. .458
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 89.8%; Score 419.2; DB 6; Length 458;
Best Local Similarity 97.3%; Pred. No. 2.6e-100;
Matches 435; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
Qy 22 CCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTTAAA 81
Db 458 CCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTTAAA 399
Qy 82 GGCT-CATGCTGGAGTGGGACTGGTGGTGGCCAGAAAGTCTCTTCCCACTGACGC 140
Db 398 GGCTCCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAAGTCTCTTCCCACTGACGC 339
Qy 141 CCCCATCAGGAGTGGGCTCTTTCCCTTCTCTTCTGCTGCTCTCTCTCTCTCTCTCT 200
Db 338 CCCCATCAGGAGTGGGCTCTTTCCCTTCTCTTCTGCTGCTGCTGCTGCTGCTGCT 279
Qy 201 TGCCATGACCTGCAGCCAGCCAGCCCGCTGGGAGAGGAGAAAGTGGGGATGGCTA 260
Db 278 TGCCATGACCTGCAGCCAGCCAGCCCGCTGGGAGAGGAGAAAGTGGGGATGGCTA 219
Qy 261 AGAAGCTGGGAGATAGGAAACAGAGAGGAGTGGGTGGGCTAGGGGGGCTGCTTAT 320
Db 218 AGAAGCTGGGAGATAGGAAACAGAGAGGAGTGGGTGGGCTAGGGGGGCTGCTTAT 159
Qy 321 TTAAGTGGTGTGTTATGATCTTATTAATTTATAAAGATATAAGGCCCTGTTCA 380
Db 158 TTAAGTGGTGTGTTATGATCTTATTAATTTATAAAGATATAAGGCCCTGTTCA 99
Qy 381 TTAAGATTTGTTCCCTTCCCTGTTCAATGTTTAAAGATTTGTTCTGTGTAATAT 440
Db 98 TTAAGAAATTTGTTCCCTTCCCTGTTCAATGTTTAAAGATTTGTTCTGTGTAATAT 39
Qy 441 GTCTTTATAATAAACAGTTAAAGCTG 467
Db 38 GTCTTTATAATAAACAGTTAAAGCTG 12

RESULT 13
AX874329/c
LOCUS AX874329 403 bp DNA linear PAT 17-DEC-2003

Db 98 TTAAGAAATTTGTTCCCTTCCCTGTTCAATGTTTAAAGATTTGTTCTGTGTAATAT 39
Qy 441 GTCTTTATAATAAACAGTTAAAGCTG 467
Db 38 GTCTTTATAATAAACAGTTAAAGCTG 12

RESULT 12
AX332852/c
LOCUS AX332852 458 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3361 from Patent WO0194629.
ACCESSION AX332852
VERSION AX332852.1 GI:18123486
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 3361 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source Location/Qualifiers
1. .458
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 89.8%; Score 419.2; DB 6; Length 458;
Best Local Similarity 97.3%; Pred. No. 2.6e-100;
Matches 435; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
Qy 22 CCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTTAAA 81
Db 458 CCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTTAAA 399
Qy 82 GGCT-CATGCTGGAGTGGGACTGGTGGTGGCCAGAAAGTCTCTTCCCACTGACGC 140
Db 398 GGCTCCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAAGTCTCTTCCCACTGACGC 339
Qy 141 CCCCATCAGGAGTGGGCTCTTTCCCTTCTCTTCTGCTGCTCTCTCTCTCTCTCTCT 200
Db 338 CCCCATCAGGAGTGGGCTCTTTCCCTTCTCTTCTGCTGCTGCTGCTGCTGCTGCT 279
Qy 201 TGCCATGACCTGCAGCCAGCCAGCCCGCTGGGAGAGGAGAAAGTGGGGATGGCTA 260
Db 278 TGCCATGACCTGCAGCCAGCCAGCCCGCTGGGAGAGGAGAAAGTGGGGATGGCTA 219
Qy 261 AGAAGCTGGGAGATAGGAAACAGAGAGGAGTGGGTGGGCTAGGGGGGCTGCTTAT 320
Db 218 AGAAGCTGGGAGATAGGAAACAGAGAGGAGTGGGTGGGCTAGGGGGGCTGCTTAT 159
Qy 321 TTAAGTGGTGTGTTATGATCTTATTAATTTATAAAGATATAAGGCCCTGTTCA 380
Db 158 TTAAGTGGTGTGTTATGATCTTATTAATTTATAAAGATATAAGGCCCTGTTCA 99
Qy 441 GTCTTTATAATAAACAGTTAAAGCTG 467
Db 38 GTCTTTATAATAAACAGTTAAAGCTG 12

RESULT 13
AX874329/c
LOCUS AX874329 403 bp DNA linear PAT 17-DEC-2003

```

DEFINITION Sequence 9234 from Patent EP1074617.
ACCESSION AX874329
VERSION AX874329.1 GI:40029100
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
TITLE Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
JOURNAL Primers for synthesizing full-length cDNA and their use
Patent: EP 1074617-A 9234 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
    source
    Location/Qualifiers
    1..403
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"
ORIGIN
Query Match 83.8%; Score 391.2; DB 6; Length 403;
Best Local Similarity 97.5%; Pred. No. 7.1e-93;
Matches 393; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 65 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGCTCGTGCCCGAGAAAGTCT 124
DB 403 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGCTCGTGCCCGAGAAAGTCT 344
QY 125 CTTCTGCCACTGACGCCGCCATCAGGAGTGGGCTCTTTTCCCGCTTCTTCTTCTGTC 184
DB 343 TTTTTCGCCACTGACGCCGCCATCAGGAGTGGGCTCTTTTCCCGCTTCTTCTTCTGTC 284
QY 185 TCCTGCTCATCGCCCTGCCATGACCTGCAGCAAGCCAGCCCGCTGGGGAGGGGAGA 244
DB 283 TCCTGCTCATCGCCCTGCCATGACCTGCAGCAAGCCAGCCCGCTGGGGAGGGGAGA 224
QY 245 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGCTAGTGGGTGGCT 304
DB 223 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGCTAGTGGGTGGCT 164
QY 305 AGGGGGCTGCTTAAAGTGGTGTATGATTTATGATTTCTTATCTAATTTATACAAAGAT 364
DB 163 AGGGGGCTGCTTAAAGTGGTGTATGATTTATGATTTCTTATCTAATTTATACAAAGAT 104
QY 365 ATTAAGCCCTGTTCAATAGAAATTTGTCCTTCCCTGCTGTTCAATTTGTTAAAGAT 424
DB 103 ATTAAGCCCTGTTCAATAGAAATTTGTCCTTCCCTGCTGTTCAATTTGTTAAAGAT 44
QY 425 TGTTCCTGTAATATGCTTTTATATAAAGCTTAAAGCTG 467
DB 43 TGTTCCTGTAATATGCTTTTATATAAAGCTTAAAGCTG 1

RESULT 14
BD154391/c
LOCUS BD154391 403 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD154391
VERSION BD154391.1 GI:27860149
KEYWORDS JP 2002191363-A/9234.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
TITLE Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
JOURNAL Primers for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 9234 09-JUL-2002;
HELIIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/9234

```

```

PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68/C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof PH Key
Location/Qualifiers
FT source 1..403
FT Location/Qualifiers
    source
    Location/Qualifiers
    1..403
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
ORIGIN
Query Match 83.8%; Score 391.2; DB 6; Length 403;
Best Local Similarity 97.5%; Pred. No. 7.1e-93;
Matches 393; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 65 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGCTCGTGCCCGAGAAAGTCT 124
DB 403 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGCTCGTGCCCGAGAAAGTCT 344
QY 125 CTTCTGCCACTGACGCCGCCATCAGGAGTGGGCTCTTTTCCCGCTTCTTCTTCTGTC 184
DB 343 TTTTTCGCCACTGACGCCGCCATCAGGAGTGGGCTCTTTTCCCGCTTCTTCTTCTGTC 284
QY 185 TCCTGCTCATCGCCCTGCCATGACCTGCAGCAAGCCAGCCCGCTGGGGAGGGGAGA 244
DB 283 TCCTGCTCATCGCCCTGCCATGACCTGCAGCAAGCCAGCCCGCTGGGGAGGGGAGA 224
QY 245 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGCTAGTGGGTGGCT 304
DB 223 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGCTAGTGGGTGGCT 164
QY 305 AGGGGGCTGCTTAAAGTGGTGTATGATTTATGATTTCTTATCTAATTTATACAAAGAT 364
DB 163 AGGGGGCTGCTTAAAGTGGTGTATGATTTATGATTTCTTATCTAATTTATACAAAGAT 104
QY 365 ATTAAGCCCTGTTCAATAGAAATTTGTCCTTCCCTGCTGTTCAATTTGTTAAAGAT 424
DB 103 ATTAAGCCCTGTTCAATAGAAATTTGTCCTTCCCTGCTGTTCAATTTGTTAAAGAT 44
QY 425 TGTTCCTGTAATATGCTTTTATATAAAGCTTAAAGCTG 467
DB 43 TGTTCCTGTAATATGCTTTTATATAAAGCTTAAAGCTG 1

RESULT 15
AC127934
LOCUS AC127934 171310 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-32806, WORKING DRAFT SEQUENCE.
ACCESSION AC127934
VERSION AC127934.3 GI:25075008
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
Ota, T., Isogai, T., Metzker, M., Lee, S., Adams, C., Alder, J.,
TITLE Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
ANYALEBECHI, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
BALDWIN, D., BANDARANAIKE, D., BARBER, M., BARNSTEAD, M., BENAHMED, F.,
BISWALO, K., BLAIR, J., BLANKENBURG, K., BLYTH, P., BROWN, M.,
BRYANT, N., BUHAY, C., BURCH, P., BURRELL, K., CALDERON, E.,

```


Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyile, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabris, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, B., Jiang, H., Johnson, B., Johnson, K., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C.E., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louissegh, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, J., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 171310)
Worley, K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 171310)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23915370.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAAV
Center clone name: CH230-32806
----- Summary Statistics

Assembly program: Phrap, version 0.990329
Consensus quality: 163795 bases at least Q40
Consensus quality: 165400 bases at least Q30
Consensus quality: 166622 bases at least Q20
Estimated insert size: 169296; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 171310: contig of 171310 bp in length.

FEATURES

source

1. 171310
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-32806"

misc_feature

1..1280
/note="wgs end extension
clone end:Sp6"

misc_feature

1567..3587
/note="wgs end extension
clone end:Sp6"

misc_feature

5295..6132
/note="clone boundary
clone end:Sp6
site:
end sequence:BZ164045"

misc_feature

168396..169235
/note="clone boundary
clone end:T7
site:
end sequence:BZ164044"

misc_feature

169684..171310
/note="wgs end extension
clone end:T7"

ORIGIN

Query Match 32.5%; Score 151.8; DB 2; Length 171310;
Best Local Similarity 63.3%; Pred. No. 3.3e-29;
Matches 307; Conservative 0; Mismatches 157; Indels 21; Gaps 4;
QY 4 TGTTCCTTATCCACCTCCGCCAGCCCTTTACTATGCTGCTGTAAGCTGCTGCTG 63
Db 116761 TGATTCCTTATTACCCCGTCCCAAAATCTTCCATCTGCTGCTGTAATGCTGCTG 116820
QY 64 CTGCTGCTGCTGCTTAAAGCTCATCTTGGAGTGGGACTGCTGGTCCCAAGAAATC 123
Db 116821 CTGCTGCCCTTCAAGGCCCACTGGGAGCTGGGCCAGTCATAGATCCTCCTT 116880
QY 124 TCTTTCGCCATGACGCCGCCCATCAGGATTCGGCCCTTCTTCCCTTC----- 173
Db 116881 GGGCTTGAAGTGGGGCTGCTGGTGGTCCCTTCTTACTGCTGGTACCCCTGCCAGGACCA 116940
QY 174 -----CTTTCTGTGCTCCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTG 227
Db 116941 GGCCTTCTGCTGCTTGTGTTGTTTGTACCTCTTGCCTTATCAGCCTGCCCATGCCAATC 117000
QY 228 CCGTGGGGAAGGGGA--GAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGA 285

```

Db      117001 CCACAGGGAAGGAGGAGGAGGATGTGGCTGAGAAAGAGAGATAGAAACAGAA 117060
QY      286 AGAGGAGTAGTGGGTGGGCTAGGGGGCTGCTTATTTAAAGTGTTGTTTATGATTCCTTA 345
Db      117061 GAGGGGAGAGTAATGGACCCAGTGGGCTGTCTTATTTAAAGTGTTGTTGATGATTCCTTA 117120
QY      346 TACTAATTTATACAAAGATATTAGGCCCTGTTTCATT-AAGAAATTTGTTCCCTTCCCTCG 404
Db      117121 TACTAATTTATAGAGATATTAGGCCCTGTTGAGTTAAAGAACTGCTCATCCCGGT 117180
QY      405 TGTTCAT--ATGTTTGTAAAGATTTCTGTTGTTAAATATGTTCTTTATATATAACAGTTTAAA 462
Db      117181 TGTTCATATGTTTGTAAAAAATGTTCCATGTAAATATGTTCTTTATATATAACAGTTTAAA 117240
QY      463 AGCTG 467
Db      117241 AGCTG 117245

RESULT 16
AC136814
LOCUS   AC136814
DEFINITION Rattus norvegicus clone CH230-179117, *** SEQUENCING IN PROGRESS
ACCESSION AC136814
VERSION   AC136814.2
KEYWORDS  HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM  Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 208810)
AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
          Allen,C., Allen,H., Altsbrooks,S., Amin,A., Anguano,D.,
          Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
          Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
          Biswal,K., Blair,J., Blaukenburg,K., Blyth,P., Brown,M.,
          Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
          Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
          Chacko,J., Chavez,D., Chen,G., Chen,K., Chen,R., Chen,Z., Chu,J.,
          Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
          Davila,M., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
          Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
          Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
          Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
          Fernandez,S., Finley,M., Flegg,N., Forbes,L., Foster,M., Foster,P.,
          Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
          Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
          Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
          Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
          Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
          Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
          Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
          Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
          Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
          Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
          Lorensuhewa,L., Louleseg,H., Lozado,R.J., Lu,X., Ma,J.,
          Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
          Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
          Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
          Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
          Morgan,M., Morris,K., Morris,S., Munidas,M., Murphy,M., Nair,L.,
          Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
          Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
          Pasternak,S., Paul,H., Perez,A., Perez,L., Franknuch,C.,
          Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
          Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
          Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
          Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
          Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
          Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

```

```

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 208810)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (08-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 208810)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24796687.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCOW
Center clone name: CH230-179117
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 192829 bases at least Q40
Consensus quality: 195654 bases at least Q30
Consensus quality: 197508 bases at least Q20
Estimated insert size: 201423; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 208810: contig of 208810 bp in length.
Location/Qualifiers
1..208810
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-179117"
complement(588..1486)
/note="clone_boundary
misc_feature

```

```

clone_end:sp6
site:EcoRI
end sequence:BH303035"
misc_feature
161506..163638
/notes="wgs_contig"
misc_feature
176821..176860
/notes="clone_boundary"
clone_end:T7
site:EcoRI
end sequence:BH303034"
misc_feature
178559..181095
/notes="wgs_end extension"
clone_end:T7"

ORIGIN
Query Match      32.5%; Score 151.8; DB 2; Length 208810;
Best Local Similarity 63.34; Pred. No. 3.3e-25;
Matches 307; Conservative 0; Mismatches 157; Indels 21; Gaps 4;

QY 4 TGTTCGCTATCCACCTCCCCCAAGCCCTTTTACCTATGCTGCTGCTAAAGCTGCTGCTG 63
   |||||
Db 34766 TGAATGCTATTACCGCTCCCAAAATCTTCCCATCTGCTGCTAAATGCTGCTGCTG 34825

QY 64 CTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGGACTGCTGCTGCTGCTGCTGCTG 123
   |||||
Db 34826 CTGCTGCTGCTTCAAGGCCACCCCTGGGAGCTGGGCCAGTCTAGAGTCTCTCACTCCTT 34885

QY 124 TCYTTCGCACTGACGCCCCCATCAGGGATGGGCTTCTTCCCTTC----- 173
   |||||
Db 34886 GGGCTTGAAGTGGGGGCTTCTGCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 34945

QY 174 -----CTTTCTGCTGCTCTCTCTGCTCTATCGGCTGCTGCTGCTGCTGCTGCTGCTG 227
   |||||
Db 34946 GGCCTTCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 35005

QY 228 CCGTGGGGAAGGGA--GAAAGTGGGGATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 285
   |||||
Db 35006 CCACAGGGAAGGGGAGGGAAGGAGATGTTGGCTGAGAAAGAGGAGAGATAGAAACAGAA 35065

QY 286 AGAGGGTAGTGGGTGGGCTAGGGGGGCTGCTTATTTAAAGTGGTCTGCTGCTGCTGCTG 345
   |||||
Db 35066 GAGGGGAGTGAAATGGACCCAGTGGGCTGCTGCTTATTTAAAGTGGTCTGCTGCTGCTG 35125

QY 346 TACTAATTTATACAAAGATTAAGGCCCTGCTTCATT-AGAAGATGCTTCCCTGCCCTG 404
   |||||
Db 35126 TACTAATTTATAGAGATTAAGGCCCTCTGAGTCTGAGTCTGAGTCTGCTGCTGCTGCTG 35185

QY 405 TGTTCA--ATGTTTGTAAAGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 462
   |||||
Db 35186 TGTTCACTATGTTTGTAAAAAATGTTCCATGTAATAATGCTTTTATAATAAGAGTTAAA 35245

QY 463 AGCTG 467
   |||||
Db 35246 AGCTG 35250

```

```

RESULT 17
BC020015 3143 bp mRNA linear ROD 12-NOV-2003
LOCUS
DEFINITION Mus musculus apelin, mRNA (cDNA clone MGC:27819 IMAGE:3483588),
complete cds.
ACCESSION BC020015
VERSION BC020015.1 GI:18043669
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3143)
REFERENCE
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

```

```

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Rahey, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 3143)
Strausberg, R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 35 Row: f Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7304894.
FEATURES
Location/Qualifiers
1..3143
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC:27819 IMAGE:3483588"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI CGAP Mam5"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..3143
/gene="Apelin"
/notes="synonym: Apelin"
/db_xref="LocusID:30878"
/db_xref="MGI:1353624"
329..562
/codon_start=1
/product="apelin"
/protein_id="AAH20015.1"
/db_xref="GI:18043670"
/db_xref="LocusID:30878"
/translation="MNLRLCVQALLLLSLTAVCGVPLMPPDGTGLEGSMRYLVK
PTRSRTPGAWQGRKFRQRRLSHKGMPFP"
ORIGIN

```

COMMENT

On May 2, 2002 this sequence version replaced gi_120330317.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-351G20 is from the RPCI-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6.

FEATURES

```

source
    1. 189982
        /locat:cn/Qualifiers
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="X"
        /clone="RP23-351G20"
        /clone_lib="RPCI-23"

```

ORIGIN

```

Query Match      27.8%; Score 129.8; DB 10; Length 189982;
Best Local Similarity 69.0%; Pred. No. 2.3e-23;
Matches 245; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

QY 121 GTCTCTCTGCACTGACGCCCATCAGGATGGGCTTCTTCCCTTCTTCTTCT 179
Db 1596 GCCTCTCTCTGCTGTGTACCTTACAGGACCGGCTTCTTCTTCTTCTTCT 1637

QY 180 GTGTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 239
Db 1636 GCACCTCTGCTTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1577

QY 240 GGAGAAAGTGGGATGCTGAAGAAAGTGGGATGAGGACAGAGAGTGTGGGT 299
Db 1576 T-----TGCTGAGAAAGAGGAGATAGA-AGCACAGAGGGGAGTAAAC 1532

QY 300 GGGCTAGGGGGCTGCTTATTAAGTGGTGTGTTATGATCTTATCTATCTATCA 359
Db 1531 GGGCCAGTGGCTGTCTTATTAAGTGGTGTGTTATGATCTTATCTATCTATCA 1472

QY 360 AAGATATTAAGCCCTGTTTCATTAAGAAATGTTTCCCTTCCCTTCCCTTCCCT 412
Db 1471 GAGATATTAAGCCCTTGTAGTATAGAGAACTGTCCTTCCCTTCCCTTCCCTAT 1412

QY 413 GTTTGTAAGATTGTTCTGTTGAATATGTTTATATTAATTAACAGTTAAAGCTG 467
Db 1411 GTTTGTAAGATTGTTCTGTTGAATATGTTTATATTAATTAAGATTAAAGTTG 1357

```

```

RESULT 20
AC055817 AC055817 203805 bp DNA linear ROD 20-JAN-2003
LOCUS Mus musculus, clone RP23-15417, complete sequence.
DEFINITION AC055817
ACCESSION AC055817
VERSION AC055817.12 GI:27804302
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

1 (bases 1 to 203805)

Birren, B., Nusbaum, C. and Lander, E.

AUTHORS

Mus musculus, clone RP23-15417

TITLE

Unpublished

JOURNAL

2 (bases 1 to 203805)

REFERENCE

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lanazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lie, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPhee, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 203805)

REFERENCE

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,

Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gort, S., Graham, L., Grand-Pierre, N., Hafez, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,

Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,

Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,

Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,

Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,

Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (10-DEC-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 203805)

REFERENCE

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,

Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,

Collamore, A., Cooke, A., Cooke, P., Corum, B., Dearellano, K.,

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,

Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,

Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,

Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

Lindblad-Toh, K., Liu, G., Lu, A., Mabbitt, R., MacLean, C.,

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,

Meldrim, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J.,

O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,

Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,

Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,

Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (20-JAN-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jan 20, 2003 this sequence version replaced gi:26225143.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL

COMMENT

FLI CDNA.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 598)
Woessner J., Tan, F., Marra, M., Kucaba, T., Yandell, M., Martin, J.,
Marth, G., Bowers, L., Wylie, T., Bowers, Y., Steptoe, M., Theising, B.,
Geisel, S., Allen, M., Underwood, K., Chappell, J., Person, B.,
Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,
Schur, R., Ritter, E., Kohn, S., Swaller, T., Behrmer, K., Hillier, L.,
Wilson, R. and Waterston, R.
Full Clone Sequencing of the Longest Available Member from Each
Unigene Cluster
Unpublished
2 (bases 1 to 598)
Waterston, R.
Direct Submission
Submitted (24-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES
source

ORIGIN

Query Match 27.2%; Score 127; DB 9; Length 598;
Best Local Similarity 100.0%; Pred. No. 1e-22;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 TCTTATCTAATTTATACAAAGATTAAAGCCCTGTTTCATTAAAGAAATTTCCCTTC 400
|||||
DB 1 TCTTATCTAATTTATACAAAGATTAAAGCCCTGTTTCATTAAAGAAATTTCCCTTC 60
|||||

QY 401 CCTGTGTTCAATGTTGTAAGATGTTCTGTGTAATATGCTTTATATAAACAAGTTA 460
DB 61 CCTGTGTTCAATGTTGTAAGATGTTCTGTGTAATATGCTTTATATAAACAAGTTA 120
|||||

QY 461 AAAGCTG 467
|||||

DB 121 AAAGCTG 127
|||||

RESULT 22
AX211242
LOCUS AX211242 255 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 884 from Patent WO0157058.
ACCESSION AX211242
VERSION AX211242.1 GI:15425502

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL

FEATURES
source

ORIGIN

Query Match 15.9%; Score 74.2; DB 6; Length 255;
Best Local Similarity 66.9%; Pred. No. 1e-08;
Matches 168; Conservative 8; Mismatches 66; Indels 9; Gaps 5;

QY 199 CCTGCCATGACCTGCAGCCCAAGCCCGCCCGTGGGGAAGGGGA--GAAAGTGGGGATG 256
|||||
DB 5 CTTCGCTTATCAGCCTGCGCAATCCACAGGGAACSSGAGGGAAGGAGATGTTG 64
|||||

QY 257 GCTAAGAAAGCTGGGAGATAGGGAAACAGAGAGAGGTAGTGGCTAGGGGGGCTGCC 316
|||||
DB 65 GCTGASAAASMSGAGAGATASAVACAGAGAGGGGGAGTGAATGGACCCGCTGTC 124
|||||

QY 317 TTATTT-AAAGTGGTCTTATGATCTTATACATA-ATTATACAAAGATATTAAAGCCC 374
|||||
DB 125 TTATTTCAAGTGGTGTGTATGATCTTATACATCTATATAGAGATATTAAAGCCC 184
|||||

QY 375 TGTTCATT-AAAGAAATTTGTCCTTCCCTG---TGTTCAATGTTTGAAGATTTGTC 429
DB 185 TCTGAGTTAAGAAACTSYCTCATCCCGTCTCTCACTCATCTTTGTAATAATTTGTC 244
|||||

QY 430 TGTGTAATAT 440
DB 245 CATGCTAACAT 255
|||||

RESULT 23
AX344562 73778 bp DNA linear PAT 01-FEB-2002
LOCUS AX344562
DEFINITION Sequence 13 from Patent WO0200932.
ACCESSION AX344562
VERSION AX344562.1 GI:18492448
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source

Olek, A., Piepenbrock, C. and Berlin, K.
Diagnosis of known genetic parameters within the mh
Patent: WO 0200932-A 13 03-JAN-2002;
Epigenomics AG (DE)
Location/Qualifiers
1. 73778
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length

of seq 2: 3.673778 <223>--split as follows--seq 14
 0.000.001 TO 0.349.980--seq 15 0.300.001 649.980--seq 16
 600.001 949.980--seq 17 900.001 1.249.980--seq 18 1.200.001
 1.549.980--seq 19 1.500.001 1.849.980--seq 20 1.800.001
 2.149.980--seq 21 2.100.001 2.449.980--seq 22 2.400.001
 2.749.980--seq 23 2.700.001 3.049.980--seq 24 3.000.001
 3.349.980--seq 25 3.300.001 3.649.980--seq 26 3.600.001
 3.673.778"

ORIGIN

Query Match 9.9%; Score 46.4; DB 6; Length 73778;
 Best Local Similarity 50.0%; Pred. No. 0.31;
 Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 230 GTGGGAAGGGAGAAAGTGGGGGATGGCTATGAAGAAGCTGGAGATAGGAAACAGAAGAG 289
 DB 42743 GAGTGGAAAGTTTGGAGGGGAGGGGTTTCTAATTTATTTTGAAGATTTGTGATTGAAAAGG 42802

QY 290 GGTAGTGGCTGGCTAGGGGGCTGCCTTATTAAAGTGGTCTTTATGATCTTATATCT 349
 DB 42803 ATTGAGGAATGGGTAAAGAGGATCGTTTTTTTATAGTATTCATTAAGAAATTTCCCTTCCCTGTGTTTC 42862

QY 350 AATTATACAAGATATTAAGGCCCTGTTCATTAAAGAAATTTGTCCTTCCCTGTGTTTC 409
 DB 42863 TTTTGTGTTTATTTTATTTAGGAGAGATTAAGAAGGAAATGTTTGTGTTT 42922

QY 410 AATGTTTGAAGATGTTCTGTGTAATAATGCTTTATATAAACAAGTTAA 461
 DB 42923 TATGTCGAGTGTGTTTTTTTGTGGAGTTTTTATTTGGAGATGGAGTATTTA 42974

RESULT 24

AX344561
 LOCUS AX344561 349980 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 12 from Patent WO200932.
 ACCESSION AX344561
 VERSION AX344561.1 GI:18492447
 KEYWORDS synthetic construct
 SOURCE synthetic construct
 ORGANISM artificial sequences.

REFERENCE 1
 Olek,A., Piepenbrock,C. and Berlin,K.
 Diagnosis of Known genetic parameters within the mhc
 Patent: WO 020932-A 12 03-JAN-2002;
 Epigenomics AG (DE)
 Location/Qualifiers
 1..349980
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>--split as follows--seq 01 0.000.001 TO 0.349.980--seq 02 0.300.001 649.980--seq 03 600.001 949.980--seq 04 900.001 1.249.980--seq 05 1.200.001 1.549.980--seq 06 1.500.001 1.849.980--seq 07 1.800.001 2.149.980--seq 08 2.100.001 2.449.980--seq 09 2.400.001 2.749.980--seq 10 2.700.001 3.049.980--seq 11 3.000.001 3.349.980--seq 12 3.300.001 3.649.980--seq 13 3.600.001 3.673.778 <223>--Original length of seq 2: 3.673778 <223>--split as follows--seq 14 0.000.001 TO 0.349.980--seq 15 0.300.001 649.980--seq 16 600.001 949.980--seq 17 900.001 1.249.980--seq 18 1.200.001 1.549.980--seq 19 1.500.001 1.849.980--seq 20 1.800.001 2.149.980--seq 21 2.100.001 2.449.980--seq 22 2.400.001 2.749.980--seq 23 2.700.001 3.049.980--seq 24 3.000.001 3.349.980--seq 25 3.300.001 3.649.980--seq 26 3.600.001 3.673.778"

FEATURES
source

Query Match 9.9%; Score 46.4; DB 6; Length 349980;
 Best Local Similarity 50.0%; Pred. No. 0.33;
 Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

ORIGIN

QY 230 GTGGGAAGGGAGAAAGTGGGGGATGGCTATGAAGAAGCTGGAGATAGGAAACAGAAGAG 289
 DB 42743 GAGTGGAAAGTTTGGAGGGGAGGGGTTTCTAATTTATTTTGAAGATTTGTGATTGAAAAGG 42802

QY 290 GGTAGTGGCTGGCTAGGGGGCTGCCTTATTAAAGTGGTCTTTATGATCTTATATCT 349
 DB 42803 ATTGAGGAATGGGTAAAGAGGATCGTTTTTTTATAGTATTCATTAAGAAATTTCCCTTCCCTGTGTTTC 42862

QY 350 AATTATACAAGATATTAAGGCCCTGTTCATTAAAGAAATTTGTCCTTCCCTGTGTTTC 409
 DB 42863 TTTTGTGTTTATTTTATTTAGGAGAGATTAAGAAGGAAATGTTTGTGTTT 42922

QY 410 AATGTTTGAAGATGTTCTGTGTAATAATGCTTTATATAAACAAGTTAA 461
 DB 42923 TATGTCGAGTGTGTTTTTTTGTGGAGTTTTTATTTGGAGATGGAGTATTTA 42974

QY 230 GTGGGAAGGGAGAAAGTGGGGGATGGCTATGAAGAAGCTGGAGATAGGAAACAGAAGAG 289
 DB 342743 GAGTGGAAAGTTTGGAGGGGAGGGGTTTCTAATTTATTTTGAAGATTTGTGATTGAAAAGG 342802

QY 290 GGTAGTGGCTGGCTAGGGGGCTGCCTTATTAAAGTGGTCTTTATGATCTTATATCT 349
 DB 342803 ATTGAGGAATGGGTAAAGAGGATCGTTTTTTTATAGTATTCATTAAGAAATTTCCCTTCCCTGTGTTTC 342862

QY 350 AATTATACAAGATATTAAGGCCCTGTTCATTAAAGAAATTTCCCTTCCCTGTGTTTC 409
 DB 342863 TTTTGTGTTTATTTTATTTAGGAGAGATTAAGAAGGAAATGTTTGTGTTT 342922

QY 410 AATGTTTGAAGATGTTCTGTGTAATAATGCTTTATATAAACAAGTTAA 461
 DB 342923 TATGTCGAGTGTGTTTTTTTGTGGAGTTTTTATTTGGAGATGGAGTATTTA 342974

RESULT 25

AX344561
 LOCUS AX344561 134971 bp DNA linear HTG 31-OCT-2002
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 11 clone OSJNBa0059H21, *** SEQUENCING IN PROGRESS ***, 5 ordered pieces.
 ACCESSION AX344561
 VERSION AC116367.9 GI:24431629
 KEYWORDS HTG; HTGS PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.
 1 (bases 1 to 134971)
 Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M., Overton II,L., Taitrin,T., Kim,M., Bera,J., Jin,S., Fadrosch,D.W., Tallon,L., Koo,H., Zismann,V., Hsiao,J., Blunt,S., Vanaken,S., Riedmuller,S.B., Utterbach,T., Feldhlyum,T., Yang,Q., Haas,B., Suh,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and Fraser,C.
 Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0059H21 BAC genomic sequence

TITLE
sequence

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 134971)
 AUTHORS Buell,R.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 REFERENCE 3 (bases 1 to 134971)
 AUTHORS Buell,R.
 JOURNAL Direct Submission

COMMENT

Submitted (31-OCT-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 On Oct 31, 2002 this sequence version replaced gi:23308041.
 * NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and the accession number will be preserved.
 * 1 29814: contig of 29813 bp in length
 * 29814 29813: gap of unknown length
 * 29914 53824: contig of 23911 bp in length
 * 53825 53924: gap of unknown length
 * 53925 71255: contig of 17331 bp in length
 * 71256 71355: gap of unknown length
 * 71356 115072: contig of 43717 bp in length
 * 115073 115172: gap of unknown length
 * 115173 134971: contig of 19799 bp in length.
 Location/Qualifiers
 1..134971
 /organism="Oryza sativa (japonica cultivar-group)"

FEATURES
source


```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/cdb="CH230-402H13"
1..1286
/misc_feature
note="wgs_end_extension
clone_end:T7"
1816..2986
/misc_feature
note="clone_boundary
clone_end:T7"
site:
end_sequence:B2254305"
202874..204064
/misc_feature
note="clone_boundary
clone_end:Sp6"
site:
end_sequence:B2254306"
207192..208575
/misc_feature
note="wgs_end_extension
clone_end:Sp6"
208626..211683
/misc_feature
note="wgs_end_extension
clone_end:Sp6"

```

ORIGIN

```

Query Match          9.6%; Score 45; DB 2; Length 212887;
Best Local Similarity 58.6%; Pred. No. 0.76;
Matches 78; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 4 TGTTTGCTATCCACCTCCGCCAGCCCTTACCTATGCTGCTGTAACGCTGCTGCTG 63
|||||
Db 36657 TGACAGCTGCTGCTCCCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 36598
|||||

QY 64 CTGCTGCTGCTGCTTAAAGCTCATGCTTGAGCTGGGCACTGCTGCTGCCAGAAAGTC 123
|||||
Db 36597 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36538
|||||

QY 124 TCTCTGCGCACTG 136
|||||
Db 36537 GCTTCTGCTACTG 36525
|||||

```

RESULT 27

```

AC105482
LOCUS      Rattus norvegicus clone CH230-20G17, WORKING DRAFT SEQUENCE, 13
DEFINITION
unordered pieces.
AC105482.4 GI:24818814
VERSION    HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 272030)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Aryalabechei,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,V., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flegg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

```

REFERENCE AUTHORS

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNAU
Center clone name: CH230-20G17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 255866 bases at least Q40
Consensus quality: 259060 bases at least Q30
Consensus quality: 260838 bases at least Q20
Estimated insert size: 263005; sum-of-contigs estimation table.

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenschew,L., Loulseged,H., Lozard,R.J., Lu,X., Ma,J.,
Maheehwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Mirja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Flopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Fuazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmari,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE

JOURNAL

Direct Submission
Unpublished
2 (bases 1 to 272030)

AUTHORS

Worley,K.C.

TITLE

JOURNAL

Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 9, 2002 this sequence version replaced gi:23267948.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>) Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNAU

Center clone name: CH230-20G17

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 255866 bases at least Q40

Consensus quality: 259060 bases at least Q30

Consensus quality: 260838 bases at least Q20

Estimated insert size: 263005; sum-of-contigs estimation

Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 13 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 5173: contig of 5173 bp in length
 5174 5273: gap of unknown length
 5274 15812: contig of 10539 bp in length
 15813 15912: gap of unknown length
 15913 15993: contig of 3681 bp in length
 15994 19694 258497: contig of 238804 bp in length
 258498 258597: gap of unknown length
 258598 259007: contig of 1210 bp in length
 259008 25907: gap of unknown length
 25909 26029: contig of 1022 bp in length
 26030 261029: gap of unknown length
 261030 262187: contig of 1158 bp in length
 262188 262287: gap of unknown length
 262288 263445: contig of 1158 bp in length
 263446 263545: gap of unknown length
 263546 264973: contig of 1427 bp in length
 264974 265072: gap of unknown length
 265073 266174: contig of 1002 bp in length
 266175 267472: gap of unknown length
 267473 267572: contig of 1298 bp in length
 267573 269822: gap of unknown length
 269828 269927: gap of unknown length
 269928 272030: contig of 2103 bp in length.

FEATURES
 source 1..272030
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clones="CH230-20617"
 1..1229
 /note="wgs end_extension
 clone_end:T7"
 misc_feature 1969..2830
 /note="clone_boundary
 clone_end:T7
 site:ECORI
 end_sequence:BH320006"
 misc_feature 15913..17329
 /note="wgs contig"
 misc_feature 237547..238337
 /note="clone_boundary
 clone_end:Sp6
 site:ECORI
 end_sequence:BH320007"

ORIGIN

Query Match 9.6%; Score 45; DB 2; Length 272030;
 Best Local Similarity 58.6%; Pred. No. 0.77;
 Matches 78; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 Qy 4 TGTTCCTATCCACCTCCGCCCGCCCTTTACTATGCTGCTGCTAAGCTGCTGCTG 63
 Db 64933 TGACAGCTGCTGCTCCCGCTACTGCTGCTCTGCTGCTGCTGCTGCTGCTG 64992
 Qy 64 CTGCTGCTGCTGCTTAAGGCTCATGCTTGAGTGGGACTGTCGCTGCCAGAAAGTC 123
 Db 64993 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 65052

Qy 124 TCCTCTGCCACTG 136
 Db 65053 GCCTCTGCTACTG 65065
 RESULT 28
 AC101159
 LOCUS AC101159 71781 bp DNA linear HTG 23-NOV-2001
 DEFINITION Mus musculus clone RP23-173D8, LOW-PASS SEQUENCE SAMPLING.
 AC101159
 VERSION AC101159.1 GI:17059933
 KEYWORDS HTG; HTGS PHASE0.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 71781)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-173D8
 Unpublished
 2 (bases 1 to 71781)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunthang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Riaback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L16056
 Center clone name: 173_D_8

 * NOTE: This record contains 86 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1 737: contig of 737 bp in length
 * 738 837: gap of 100 bp
 * 838 1555: contig of 718 bp in length
 * 1556 1655: gap of 100 bp

```
* 1656 2389: contig of 734 bp in length
* 2390 2489: gap of 100 bp
* 2490 3228: contig of 739 bp in length
* 3229 3328: gap of 100 bp
* 3329 4067: contig of 739 bp in length
* 4068 4167: gap of 100 bp
* 4168 4887: contig of 720 bp in length
* 4888 4987: gap of 100 bp
* 4989 5716: contig of 728 bp in length
* 5717 5815: gap of 100 bp
* 5816 6559: contig of 744 bp in length
* 6560 7403: gap of 100 bp
* 7404 7503: gap of 100 bp
* 7504 8249: contig of 746 bp in length
* 8250 9071: gap of 100 bp
* 9072 9171: contig of 722 bp in length
* 9172 9906: contig of 735 bp in length
* 9907 10006: gap of 100 bp
* 10007 10749: contig of 743 bp in length
* 10750 10849: gap of 100 bp
* 10850 11601: contig of 751 bp in length
* 11602 11701: gap of 100 bp
* 11702 12457: contig of 757 bp in length
* 12458 12557: gap of 100 bp
* 12558 13307: contig of 750 bp in length
* 13308 14123: contig of 722 bp in length
* 14124 14229: gap of 100 bp
* 14230 14935: contig of 706 bp in length
* 14936 15035: gap of 100 bp
* 15036 15768: contig of 733 bp in length
* 15769 15868: gap of 100 bp
* 15869 16604: contig of 736 bp in length
* 16605 16704: gap of 100 bp
* 16705 17439: contig of 735 bp in length
* 17440 17539: gap of 100 bp
* 17540 18275: contig of 736 bp in length
* 18276 18376: gap of 100 bp
* 18377 19100: contig of 725 bp in length
* 19101 19200: gap of 100 bp
* 19201 19938: contig of 738 bp in length
* 19939 20038: gap of 100 bp
* 20039 20888: contig of 750 bp in length
* 20889 21630: contig of 742 bp in length
* 21631 21730: gap of 100 bp
* 21731 22457: contig of 727 bp in length
* 22458 22557: gap of 100 bp
* 22559 23308: contig of 751 bp in length
* 23309 24146: gap of 100 bp
* 24147 24246: gap of 100 bp
* 24247 24971: contig of 725 bp in length
* 24972 25071: gap of 100 bp
* 25072 25816: contig of 745 bp in length
* 25817 25916: gap of 100 bp
* 25917 26659: contig of 743 bp in length
* 26660 26759: gap of 100 bp
* 26760 27504: contig of 745 bp in length
* 27505 27604: gap of 100 bp
* 27605 28376: contig of 772 bp in length
* 28377 28476: gap of 100 bp
* 28477 29212: contig of 736 bp in length
* 29213 30041: contig of 729 bp in length
* 30042 30141: gap of 100 bp
* 30142 30874: contig of 732 bp in length
* 30875 31708: gap of 100 bp
* 30974 31709: contig of 735 bp in length
* 31710 32507: gap of 100 bp
* 32507: contig of 699 bp in length
```

```
* 32508 32607: gap of 100 bp
* 32608 33340: contig of 733 bp in length
* 33341 34150: gap of 100 bp
* 34151 34251: gap of 100 bp
* 34252 34998: contig of 748 bp in length
* 34999 35098: gap of 100 bp
* 35099 35845: contig of 747 bp in length
* 35846 35945: gap of 100 bp
* 35946 36690: contig of 745 bp in length
* 36691 37531: gap of 100 bp
* 37532 37632: gap of 100 bp
* 37633 38368: contig of 737 bp in length
* 38369 38468: gap of 100 bp
* 38469 39124: contig of 656 bp in length
* 39125 39224: gap of 100 bp
* 39225 39964: contig of 740 bp in length
* 39965 40064: gap of 100 bp
* 40065 40798: contig of 734 bp in length
* 40799 40898: gap of 100 bp
* 40899 41647: contig of 749 bp in length
* 41648 41747: gap of 100 bp
* 41748 42492: contig of 745 bp in length
* 42493 42592: gap of 100 bp
* 42593 43313: contig of 721 bp in length
* 43314 43413: gap of 100 bp
* 43414 44157: contig of 744 bp in length
* 44158 44257: gap of 100 bp
* 44258 45006: contig of 749 bp in length
* 45007 45106: gap of 100 bp
* 45107 45864: contig of 758 bp in length
* 45865 45964: gap of 100 bp
* 45965 46713: contig of 749 bp in length
* 46714 46813: gap of 100 bp
* 46814 47557: contig of 744 bp in length
* 47558 47657: gap of 100 bp
* 47658 48384: contig of 727 bp in length
* 48385 48484: gap of 100 bp
* 48485 49221: contig of 737 bp in length
* 49222 49321: gap of 100 bp
* 49322 50043: contig of 722 bp in length
* 50044 50143: gap of 100 bp
* 50144 50860: contig of 737 bp in length
* 50861 50960: gap of 100 bp
* 50961 51702: contig of 722 bp in length
* 51703 51802: gap of 100 bp
* 51803 52511: contig of 709 bp in length
* 52512 52611: gap of 100 bp
* 52612 53350: contig of 739 bp in length
* 53351 53450: gap of 100 bp
* 53451 54182: contig of 732 bp in length
* 54183 54282: gap of 100 bp
* 54283 55030: contig of 748 bp in length
* 55031 55130: gap of 100 bp
* 55131 55830: contig of 700 bp in length
* 55831 55930: gap of 100 bp
* 55931 56688: contig of 758 bp in length
* 56689 56788: gap of 100 bp
* 56789 57551: contig of 763 bp in length
* 57552 57651: gap of 100 bp
```

Query Match 9.5%; Score 44.2; DB 2; Length 71781;
Best Local Similarity 61.9%; Pred. No. 1.2;
Matches 70; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2 AATGTTGCTATCCACCTCCCAAGGCCCTTACCTATGCTGCTAAAGCTGCTGC 61
|||||

DB 16221 AATGGGGGTTTCACTGTACAAAGGCCAGTTCTGCTGCTGCTGCTGC 16280
|||||

QY 62 TGCTGCTGCTGCTGCTTAAAGCTCATGCTTGAGATGGGACTGCTGGGTGCC 114
|||||

DB 16281 TGCTGCTGCTGCTGCTAAAGCAGACTAAAGTTGGAGTGGGATCTTCAGTTCC 16333
|||||

RESULT 29

AC119515
 LOCUS
 DEFINITION AC119515 213417 bp DNA linear HTG 15-NOV-2002
 Rattus norvegicus clone CH230-40602, *** SEQUENCING IN PROGRESS

ACCESSION

AC119515 GI:25012225
 HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 213417)
 Muzny, D., Marie, E., Metzker, M., Lee, J., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jarkovsky, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensheua, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwaokeme, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhauser, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 213417)

REFERENCE

Worley, K.C.
 Direct Submission
 Submitted (28-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 213417)
 Rat Genome Sequencing Consortium.
 Direct Submission

AUTHORS

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 15, 2002 this sequence version replaced gi:23616941.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GUQW

Center clone name: CH230-40602

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 197852 bases at least Q40

Consensus quality: 199838 bases at least Q30

Consensus quality: 201193 bases at least Q20

Estimated insert size: 207854; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 213417: contig of 213417 bp in length.

Location/Qualifiers

1..213417

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-40602"

1..1798

/note="wgs end extension"

clone end:Sp6"

6829..7711

/note="clone boundary"

clone end:Sp6

site:

end sequence:BZ148299"

206919..208667

/note="wgs contig"

211434..213417

/note="wgs_contig"

ORIGIN

Query Match 9.5%; Score 44.2; DB 2; Length 213417;

Best Local Similarity 50.7%; Pred. No. 1-2;

Matches 106; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY

232 GGCGAGGGGAGAGATGGGGATGCTAGAAAGTGGAGATAGGGAACAGAGGG 231

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||


```

misc_feature 237046..239139
              /note="wgs_contig"

ORIGIN
Query Match      9.5%; Score 44.2; DB 2; Length 242655;
Best Local Similarity 50.7%; Pred. No. 1.2;
Matches 106; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 232 GGGGAAGGGGAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAAGAGAGGG 291
      |||
Db 191598 GAGAGAGGAGAGAGGAGGAGAGGAAAGAGAGGAGGGGAGGAGAGAGAAAGAAAGG 191539
      |||

QY 292 TAGTGGGTGGCTAGGGGGGCTGCTTATTTAAAGTGGTCTTTATGATCTTATATACTAA 351
      |||
Db 191538 TGGTGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGTGTAGGAAATCCACTAGAA 191479
      |||

QY 352 TTTATACAAAGATATTAAGGCCCTGTTTCATTAAGAAATGTTCCTTCCCTGTGTTCAA 411
      |||
Db 191478 ATAAAGAGTAATTAAGAAAGGAGAGACTATGAATCATTTCTTGTCTTTCTTAAGGTTAT 191419
      |||

QY 412 TGTGTTGAAGATTTGTTCTGTGTAATAT 440
      |||
Db 191418 TTTTATATTTCTTTCTTTACTAAGAT 191390
      |||

RESULT 31
AL606973
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-173D8 on chromosome 4, complete
sequence.
ACCESSION AL606973
VERSION AL606973.29 GI:23497888
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Clark, G.
DIRECT SUBMISSION
Submitted (04-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueres@sanger.ac.uk
Cloned request: clonerequest@sanger.ac.uk
On Oct 3, 2002 this sequence version replaced gi:23337158.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humqueres@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/c_elegans/wormpep
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers

```

```

1..243369
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-173D8"
/clone_lib="RPI-23"

ORIGIN
Query Match      9.5%; Score 44.2; DB 10; Length 243369;
Best Local Similarity 61.9%; Pred. No. 1.2;
Matches 70; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2 AATGTTTGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTGCTGCTGCTGCTGCTG 61
      |||
Db 139975 AATGAGGGGTTTCACTGTACAAAGGCCAGTTCGTGCTGCTGCTGCTGCTGCTGCTG 140034
      |||

QY 62 TGTGCTGCTGCTGCTGCTTAAAGGCTCATGTTGGAGTGGGAGCTGTCGTTGCC 114
      |||
Db 140035 TGTGCTGCTGCTGCTAAGCAGACTAAAGTTGGACTGTGGATCTTCAGTTC 140087
      |||

RESULT 32
AL646436/c
LOCUS
DEFINITION Pan troglodytes chromosome UNK clone RP43-16B7, *** SEQUENCING IN
PROGRESS ***, 15 unordered pieces.
ACCESSION AL646436
VERSION AL646436.2 GI:34147022
KEYWORDS HTG; HTGS PHASE1.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
AUTHORS Wilton, R.K.
DIRECT SUBMISSION
Submitted (13-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 173254)
Wilton, R.K.
Direct Submission
Submitted (23-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
On Aug 23, 2003 this sequence version replaced gi:33621016.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: CPT016B07
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164865 bases at least Q40
Consensus quality: 166558 bases at least Q30
Consensus quality: 167480 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

QY 71 TGCTGCTTAAGGCTCATGCTTGGAGTGGGAGCTGGTCCGCCAGAAAGTCTTCTG 130
 |||||
 Db 276 GGTCTGCTGACCTCAGGTCAGATGGTCCGATGGTCCATATCAGCTGGCTCTGTG 217
 |||||
 QY 131 CCATCGAGCGCCCAATCAGGAGTGGGCTCTTTCCGCCCTTCCCTTCTGTCGTCCTG 190
 |||||
 Db 216 TCCTGGTCTTGTCTGCGGGTCTTGTGTGCGGGTCCATATCCTCTGTCGTCGTCG 157
 |||||
 QY 191 CTATCGGCGCTGCCNAGCACTGACGCCAGCCAG 225
 |||||
 Db 156 TCTGCGCGCGCGGTGATGACCTCCAGGTCCAG 122
 |||||

RESULT 34 AC133581/C

LOCUS AC133581 155665 bp DNA linear HTG 17-DEC-2003
 DEFINITION Mus musculus chromosome 1 clone RP24-395P3 map 1, *** SEQUENCING IN
 PROGRESS ***, 6 unordered pieces.

AC133581
 VERSION AC133581.4 GI:39979530
 KEYWORDS HTG; HTGS PHASE1; HTGS FULLTOP; HTGS ACTIVEFIN.
 SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 155665)

Birren,B., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 155665)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Farko,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (14-SEP-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 155665)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeRellano,K., Faro,S.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Farko,S.,
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hagos,B., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE JOURNAL

COMMENT

Direct Submission
 Submitted (17-DEC-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 17, 2003 this sequence version replaced gi:28933642.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WtBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27198

Center clone name: 395_P_3

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 12796: contig of 12796 bp in length
 * 12797 12896: gap of 100 bp
 * 12897 59863: contig of 46967 bp in length
 * 59864 59963: gap of 100 bp
 * 59964 93811: contig of 33848 bp in length
 * 93812 93911: gap of 100 bp
 * 93912 14271: contig of 20360 bp in length
 * 14272 14371: gap of 100 bp
 * 14372 130845: contig of 16474 bp in length
 * 130846 130945: gap of 100 bp
 * 130946 155665: contig of 24720 bp in length.

FEATURES source

1. 155665
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="1"
 /map="1"
 /clone="RP24-395P3"
 /clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match 9.4%; Score 43.8; DB 2; Length 155665;
 Best Local Similarity 76.1%; Pred. No. 1.6;
 Matches 54; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 7 TTGCTATTCACCTCCGCCAAGCCCTTTACCTATGCTGTGTAAGCTGCTGCTG 66
 |||||
 Db 154263 TTCCAGTCCGCCCTCCCAACCGCGCAGCCTCTGCTGCTGCTGCTGCTG 154204
 |||||

QY 67 CTGCTGCTGCT 77

Db 154203 CTGCTGCTGCT 154193

RESULT 35

HSDJ989D7

LOCUS HSDJ989D7 119790 bp DNA linear PRI 10-MAR-2001
 DEFINITION Human DNA sequence from clone RPS-989D7 on chromosome 20. Contains
 the 3' end of a novel gene, STSs and GSSs, complete sequence.

ACCESSION AL121911

VERSION AL121911.15 GI:9864671

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 119790)

AUTHORS Griffiths,C.

TITLE
JOURNAL

COMMENT
Direct Submission
Submitted (02-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi.9714378.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone
RP5-989D7. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP5-989D7 is at 1 in this sequence. The
true left end of clone RP5-859D4 is at 119691 in this sequence. The
true right end of clone RP11-199014 is at 19977 in this sequence.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP5-989D7 is from the
library RPCI-5 constructed by the group of Pieter de Jong. For
further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

FEATURES
source
1..11970
/location=Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-989D7"
/clone_lib="RPCI-5"
1..161
/gene="dj989D7.1"
1..161
/gene="dj989D7.1"
/product="dj989D7.1 (novel transcript)"
/notes="continued from BA199014.1 in Em:AL162504
match: cDNAs: Em:AY007089"
/evidence=not_experimental
148..153
/gene="dj989D7.1"
945..3332
/notes="TIGGER1 repeat: matches 1..2418 of consensus"
4130..4386
/note="match: GSS: Em:AQ380734"
4394..4326
/notes="LIMB8 repeat: matches 5834..5865 of consensus"
7201..8108
/notes="LIM4 repeat: matches 4784..5705 of consensus"
complement(8165..8633)
/note="match: GSS: Em:AQ277139"
8629..10044
/notes="LIM4 repeat: matches -260..1191 of consensus"
10045..10190
/notes="73 copies 2 mer tt 57% conserved"
10382..10435
/notes="27 copies 2 mer tt 74% conserved"
11843..12150
/note="match: GSS: Em:AQ429711"
/notes="AluX repeat: matches 1..308 of consensus"

repeat_region
12350..12538
/note="MIR repeat: matches 60..249 of consensus"
repeat_region
12659..13160
/note="MLT1J repeat: matches 13..513 of consensus"
repeat_region
14687..14768
/note="HAI1 repeat: matches 271..360 of consensus"
repeat_region
15106..15438
/note="MER2 repeat: matches 1..345 of consensus"
misc_feature
15368..15768
/note="match: GSS: Em:AQ772240"
misc_feature
15375..15560
/note="match: GSS: Em:AQ012918"
repeat_region
15807..16507
/note="LIMB2 repeat: matches 5469..6165 of consensus"
repeat_region
16561..17220
/note="LIMB2 repeat: matches 4871..5474 of consensus"
repeat_region
17218..17509
/note="LIMB8 repeat: matches 5848..6168 of consensus"
repeat_region
17581..17733
/note="MER8 repeat: matches 87..239 of consensus"
repeat_region
17734..17987
/note="LIMB1 repeat: matches 5782..6043 of consensus"
repeat_region
18038..18183
/note="LIM10 repeat: matches 6174..6316 of consensus"
repeat_region
18188..18525
/note="LIMB7 repeat: matches 2409..2740 of consensus"
repeat_region
18526..18878
/note="THE1A repeat: matches 1..354 of consensus"
repeat_region
18879..22631
/note="LIMB7 repeat: matches 2740..6162 of consensus"
repeat_region
23729..24007
/note="LIMB1 repeat: matches 5598..5892 of consensus"
repeat_region
24031..24127
/note="LIM10 repeat: matches 5993..6095 of consensus"
repeat_region
24128..24439
/note="MER7A repeat: matches 1..345 of consensus"
repeat_region
24440..24524
/note="LIM10 repeat: matches 6095..6317 of consensus"
repeat_region
25218..27841
/note="LIMC1 repeat: matches 3550..6326 of consensus"
complement(27586..28074)
/note="match: GSS: Em:AQ605230"
repeat_region
28283..28491
/note="LTR16A repeat: matches 148..378 of consensus"
repeat_region
28592..28729
/note="MIR repeat: matches 26..167 of consensus"
repeat_region
29416..29585
/note="MER5B repeat: matches 1..177 of consensus"
repeat_region
30385..30489
/note="MIR repeat: matches 40..153 of consensus"
repeat_region
30759..31042
/note="MER33 repeat: matches 47..323 of consensus"
repeat_region
31306..31349
/note="22 copies 2 mer tt 84% conserved"
repeat_region
31491..31825
/note="MER77 repeat: matches 38..400 of consensus"
complement(31606..32269)
/note="match: GSS: Em:AQ266996"
repeat_region
32371..32509
/note="LTR16C repeat: matches 81..221 of consensus"
repeat_region
32510..33153
/note="MER51B repeat: matches 1..617 of consensus"
repeat_region
33154..33323
/note="LTR16C repeat: matches 221..385 of consensus"
misc_feature
35252..35712
/note="match: GSS: Em:AQ700733"
misc_feature
35693..36037
/note="match: STS: Em:L16414"
repeat_region
35731..35802
/note="MLT1J repeat: matches 124..196 of consensus"
misc_feature
35934..36427
/note="match: GSS: Em:AQ429711"
repeat_region
35967..36020

```

/note="9 copies 6 mer tagata 75% conserved"
35988. .36019
repeat_region /note="13 copies 4 mer agat 100% conserved"
repeat_region 36212. .36570
misc_feature /note="LTR16A repeat: matches 57. .433 of consensus"
38224. .38413
repeat_region /note="match: GSS: Em:A2235965"
38910. .39037
repeat_region /note="MIR repeat: matches 74. .225 of consensus"
39063. .39103
repeat_region /note="L2 repeat: matches 2705. .2745 of consensus"
39217. .39290
repeat_region /note="L2 repeat: matches 2410. .2484 of consensus"
39231. .39673
repeat_region /note="MLT1A1 repeat: matches 4. .365 of consensus"
39678. .39933
repeat_region /note="L2 repeat: matches 2159. .2419 of consensus"
39960. .40082
repeat_region /note="L2 repeat: matches 1978. .2096 of consensus"
40421. .41176
repeat_region /note="L1PA16 repeat: matches 3166. .3932 of consensus"
41179. .43371
repeat_region /note="L1PA16 repeat: matches 3923. .6157 of consensus"
43715. .43783
repeat_region /note="MIR repeat: matches 14. .80 of consensus"
45011. .45066
repeat_region /note="L4 copies 4 mer tata 80% conserved"
45244. .45466
repeat_region /note="L2 repeat: matches 2168. .2393 of consensus"
45535. .45726
repeat_region /note="L2 repeat: matches 2557. .2746 of consensus"
46043. .46146
repeat_region /note="L2 repeat: matches 2602. .2710 of consensus"
46287. .46523
repeat_region /note="L2 repeat: matches 2022. .2266 of consensus"
46738. .46912
repeat_region /note="L2 repeat: matches 2541. .2710 of consensus"
47299. .47342
repeat_region /note="MIR repeat: matches 97. .139 of consensus"
47331. .47479
repeat_region /note="MIR repeat: matches 99. .256 of consensus"
complement(47625. .48021)
misc_feature

Query Match 9.3%; Score 43.2; DB 9; Length 119790;
Best Local Similarity 58.6%; Pred. No. 2.2;
Matches 75; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 324 AAGTGGTGTGTTATGATCTTATCTACTATTTATACAAAGATATTAAAGCCCTGTTTCATTA 383
Db 115477 AAGTCTTCAGAAATCCCTATAGAAGATTTATTTCTAGTAAGAACTGATGTTG 115536

QY 384 AGAAATGTCCTCCCTCCGCTGTCATGTTGTGTAAGATTTGTCGTGTAATATGTC 443
Db 115537 AGAAGTGATTCCTTACCTCTGCTGCTGCTGCTACAGCCAGTTGTGTGAGAATATGTG 115596

QY 444 TTTATAT 451
Db 115597 CTCATAAT 115604

RESULT 36
AC093112 59414 bp DNA linear PRI 20-MAR-2002
LOCUS Homo sapiens BAC clone RP11-621K10 from 2, complete sequence.
DEFINITION AC093112
ACCESSION AC093112
VERSION AC093112.4 GI:19551206
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 59414)
AUTHORS Sulston,J.E. and Waterston,R.

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 59414)
Waligorski,J. and Maupin,R.
The sequence of Homo sapiens BAC clone RP11-621K10
Unpublished (2001)
3 (bases 1 to 59414)
Waterston,R.H.
Direct Submission
Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 59414)
Waterston,R.H.
Direct Submission
Submitted (23-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 59414)
Waterston,R.
Direct Submission
Submitted (20-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 20, 2002 this sequence version replaced gi:10873939.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0621K10
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Ooeegawa,K., Moon,P.Y., Zhao,B., Frengen,B., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-157M22, 2000 bp overlap; the clone sequenced to the right is RP11-266L10, 2000 bp overlap. Actual start of this clone is at base position 79221 of RP11-157M22; actual end is at base position 33594 of RP11-266L10.

Location/Qualifiers
1. 59414
/organism="Homo sapiens"
/mol_type="genomic DNA"

FEATURES
source

REFERENCE
AUTHORS

Rattus.
1 (bases 1 to 343188)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, P., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kwis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwionu, G., Olarunpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, B., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Streimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.

TITLE
JOURNAL

REFERENCE
1 (bases 1 to 343188)
Worley, K. C.

AUTHORS
JOURNAL

Submitted (12-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 343188)
Rat Genome Sequencing Consortium.

REFERENCE
AUTHORS
JOURNAL

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23816565

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSMR
Center clone name: CH230-271E17
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 265870 bases at least Q40
Consensus quality: 268598 bases at least Q30
Consensus quality: 270491 bases at least Q20
Estimated insert size: 281218; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation table.

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 29799: contig of 29799 bp in length
29800: gap of unknown length
29900: contig of 110080 bp in length
139980: gap of unknown length
140079: contig of 13620 bp in length
153700: gap of unknown length
153700: gap of unknown length
153800: contig of 179581 bp in length
333481: gap of unknown length
333481: contig of 1056 bp in length
334637: gap of unknown length
334637: contig of 1222 bp in length
335859: gap of unknown length
337052: contig of 1093 bp in length
337052: gap of unknown length
337152: contig of 1298 bp in length
338450: gap of unknown length
338550: contig of 1245 bp in length
339795: gap of unknown length
339895: contig of 3294 bp in length.

FEATURES

source

1. 343188
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-271E17"
4552. 5701
/note="wgs_end_extension
clone_end:T7"
complement(17244..17621)
clone_end:T7
site:
end_sequence:RXAF33TUB"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature 153800..179198
/note="wgs_contig"
177457..178312
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ218560"
179249..180714
/note="wgs_end_extension
clone_end:Sp6"

ORIGIN

Query Match 9.2%; Score 43; DB 2; Length 343188;
Best Local Similarity 59.3%; Pred. No. 2.6;
Matches 73; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2 AATGTTGGCTATCACCTCCCAAGCCCCCTTACCTATGCTGTGCTTAAGCGTGCTGC 61
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 TGTCTGCTGCTGCTGTTAAAGGCTCATGCTTGGAGTGGGAGCTGCTCGGTCGCCAGAAG 121
Db TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC 135770

QY 122 TCT 124
Db | | |
QY 135709 TCT 135707
Db | | |

RESULT 38

AC096041

LOCUS AC096041.6 GI:24941294 linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-15A4, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
AC096041
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 243334)

AUTHORS Muzny,D.,Marie,E., Metzker,M.Lee,, Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,N., Blair,J., Blum,P.P., Burrell,K., Calderon,B.,
Bryant,N., Buhay,G., Burck,C., Planchenbourg,K., Blyth,P., Brown,M.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,Y., Chen,Z., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,X., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinsh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,P., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,X., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huliyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.I., Lebow,H., Levav,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Lit,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mathewine,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

* consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 199221: contig of 199221 bp in length.

FEATURES

source
 1. 199221
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-28/H17"
 1. 1558
 /note="wgs end extension
 clone_end:T7"
 4765..6815
 /note="wgs end extension
 clone_end:T7"
 7600..8500
 /note="clone boundary
 clone_end:T7"
 site:
 end sequence:BZ134792"
 192560..193460
 /note="clone boundary
 clone_end:Sp6
 site:
 end sequence:BZ134793"
 195186..196606
 /note="wgs end extension
 clone_end:Sp6"
 197746..199221
 /note="wgs end extension
 clone_end:Sp6"

ORIGIN

Query Match 9.1%; Score 42.6; DB 2; Length 199221;
 Best Local Similarity 58.1%; Pred. No. 3.3;
 Matches 75; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
 QY 4 TGTGTGCTGCTACCTCCCAAGCCCTTACCTATGCTGCTGCTGCTGCTGCTG 63
 Db 16884 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 16943
 QY 64 CTGCTGCTGCTCTTAAAGGCTCATGTTGGAGTGGGACTGCTGCTGCCAGAAAGTC 123
 Db 16944 CTGCTGCTGCTGCTCTTCCCTAGCTGATGGTCAGCAGAGTCTTCTTGACTCCTATACAAA 17003
 QY 124 TCTTCTGCC 132
 Db 17004 GCTGCTCC 17012

RESULT 44
 WMAE000663
 LOCUS WMAE000663 250611 bp DNA linear ROD 04-OCT-2002
 DEFINITION Mus musculus TCR beta locus from bases 1 to 250611 (section 1 of 3)
 of the complete sequence.
 ACCESSION AE000663 AE000522
 VERSION AE000663.1 GI:2358069
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 250611)
 Chen, F., Rowen, L., Hood, L. and Rotherberg, E.V.
 TITLE Differential transcriptional regulation of individual TCR V beta
 segments before gene rearrangement
 JOURNAL J. Immunol. 166 (3), 1771-1780 (2001)

MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

21103195
 11160223
 2 (bases 1 to 250611)
 Rowen, L., Koop, B.F., Boyesen, C., Wang, K., Ahearn, M.E., Qin, S.,
 Lee, I., Seto, J., Acharya, C., Ankner, M., Alving, W., Chen, L.,
 Paepker, B., Baskin, D., Jerome, N., Swartzell, S., Gilbert, T.,
 Faust, J., Loretz, C., Bumgarner, R. and Hood, L.
 Sequence of the mouse T cell receptor locus
 Unpublished

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE

3 (bases 1 to 250611)
 Rowen, L., Smit, A.F.A. and Hood, L.
 Comparison of the human and mouse T cell receptor beta and
 trypsinogen loci
 Unpublished

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

4 (bases 1 to 250611)
 Rowen, L.
 Direct Submission

COMMENT

Submitted (20-JUL-1997) Department of Molecular Biotechnology, Box
 357730 University of Washington, Seattle, Washington 98195, USA
 Sequencing methodology: high redundancy shotgun. Based on
 overlapping cosmids, the precision of this sequence is
 99.93. Libraries were derived from two strains but there was no
 discernable sequence variation between them. This sequence
 contains numerous internal duplications, especially in the
 tryptsinogen genes. Interspersed Repeats were identified with
 RepeatMasker (available from
 http://ftp.genome.washington.edu/RM/RepeatMasker.html) Several new
 repeats were identified by Arian Smit, and these have not yet been
 fully classified. Simple sequence repeats were identified with
 sputnik (available from
 http://serac.mbt.washington.edu/~chrisa/software/sputnik.html)
 The authors thank Christopher Nelson for his gift of cosmids from
 Dennis Loh's library.

FEATURES

Location/Qualifiers

1..250611
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="6"
 1..36170
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /clone="cosmid D23"
 /clone_lib="Dennis Loh"
 34707..76259
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="129"
 /db_xref="taxon:10090"
 /clone="cosmid M4B2"
 /clone_lib="subcloned from BAC 57B22, Genome Systems
 library"
 63060..97731
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /clone="cosmid 0742C"
 /clone_lib="Kai Wang"
 96883..131111
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /clone="cosmid 79A"
 110337..144911
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="cosmid 74A"

source

source

source

source

source

```

/clone_lib="Kai Wang"
/notes="this cosmid is chimeric. TCRB portion begins 6530
bases into the cosmid."
source
131543..170793
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="cosmid D15"
/clone_lib="Dennis Loh"
158907..190749
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="cosmid 6.3"
/clone_lib="Kai Wang"
190216..220895
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="cosmid 83.1"
/clone_lib="Kai Wang"
207816..239023
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="cosmid 47.2"
/clone_lib="Kai Wang"
235693..250611
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="cosmid 67.2"
/clone_lib="Kai Wang"
complement(1132..9834)
/genes="dopamine-beta-hydroxylase-like"
/notes="this gene is similar to human
dopamine-beta-hydroxylase by blastx."
complement(join(1134..1376,1660..1777,2666..2796,
4260..4319,4459..4555,4795..4886,5730..5896,6070..6172,
6347..6514,7236..7379,7627..7788,8276..8422,9547..9834),
/genes="dopamine-beta-hydroxylase-like"
/codon_start=1
/product="dopamine-beta-hydroxylase"
/protein_id="AAB69054.1"
/db_xref="GI:2358082"
/translation="MACVLLFRFLLLVLAFAFSQKRLGTPSLRYSRFLDPGRAVFL
RWDFYAEIITFEQLQTTQWVGIGITDRTYFVGSDLVGVGLPNGNVYFSQHLID
EDTLQDSQDAELRLTLDVSTTMRFSRPTCDPHRDITSDTMRVLAAYGPDPI
PKMSREHTFVKISFLLQMDQDADPDTI1HDLKI SNFTIPEDDTYACTFLPLPI
VSKGHIYKFPPIVVERNETVHVHVVYACNSVLTGTGCEYGSPPASLCSHVIA
GMAVGLSYQPDVGISIGTTPQWIRLEIHSNQNLPGRDITSGMLFTYSHLR
KIDMVLQGLISVFPIHPFGPAFAFUSYGLKDKFEELNGAPVSDIYISACILHTR
LAGRSLOALVRNQTQVQCKFSYDFNLQESRDLPHPVVKPGDILLIECHVQTLD
RDMFTFGASTINMCLTFFPYPRINISCMGYPDI1YVYNELGREASPNMELMV
LDNVEPTENIKKAEKACKESQQTLLKTI DEVENTGMDIPIPTPGPCLESTGG
KVEPDNTPAGFRAVPLALSQSNATLRPLPMIAVLFLQSLSCLLAMLQTGV"
2072..2215
/repeat_region
/rpt_family="B1_MM"
3146..3201
/repeat_region
/rpt_type=tandem
/rpt_unit="ATATA"
complement(5333..5482)
/repeat_region
/rpt_family="Lx9"
complement(8820..9051)
/repeat_region
/rpt_family="URRIA"
complement(11710..12162)
/repeat_region
/rpt_family="BMER10A"
13212..13327
/repeat_region
/rpt_family="MYS1_PL"

```

```

REFERENCE
AUTHORS      5 (bases 1 to 139786)
TITLE        McPherson,J.D. and Waterston,R.H.
JOURNAL      Direct Submission
COMMENT      Submitted (10-JAN-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
              6 (bases 1 to 139786)
              Wilson,R.
              Direct Submission
              Submitted (13-NOV-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Jan 10, 2003 this sequence version replaced gi:24211432.
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: submission@watson.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: M_BB0445C12

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. 139786
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="3"
/map="3"
/clone="RP24-445C12"
/clone_lib="RPCI-24"
1. 91
/rpt_family="L1"
repeat_region 845. .887
/rpt_family="L1"
repeat_region 995. .1167
/rpt_family="L1"
repeat_region 1231. .1284
/rpt_family="ERV1"
repeat_region 1454. .1600
/rpt_family="L1"
repeat_region 2273. .2774
/rpt_family="RMR15"
repeat_region 4149. .4592
/rpt_family="L1"
repeat_region 7500. .7532
/rpt_family="L1"
repeat_region 7533. .7724

```


| | |
|---------------|------------------------------------|
| repeat_region | 43641..43711 /rpt_family="B4" |
| repeat_region | 44073..44417 /rpt_family="L1" |
| repeat_region | 45186..45590 /rpt_family="L1" |
| repeat_region | 47166..47283 /rpt_family="B4" |
| repeat_region | 47284..47297 /rpt_family="Alu" |
| repeat_region | 48075..48802 /rpt_family="L1" |
| repeat_region | 48808..50361 /rpt_family="L1" |
| repeat_region | 50974..51573 /rpt_family="L1" |
| repeat_region | 52314..52453 /rpt_family="Alu" |
| repeat_region | 52587..52789 /rpt_family="B4" |
| repeat_region | 52839..53228 /rpt_family="ERVK" |
| repeat_region | 53349..53400 /rpt_family="B4" |
| repeat_region | 53406..53982 |

Query Match

Best Local Similarity 9.1%; Score 42.4; DB 10; Length 139786;

Matches 55; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY

42 GTCGCTCAACCTCCTCTGTCTGTCTGTCTAAAGCTCATGCTTGAGTGGG 101

|||||

Dd

12712 ACTGCTGCTGCTCTGTCTGTCTGTCTGTCTATTTTATTGTTAGTGG 12653

|||||

QY

102 ACTGTCGGTGGCCAG 117

|||||

Dd

12652 ACTGGACTGCAGGCG 12637

RESULT 47

AC091992

LOCUS

DEFINITION Homo sapiens chromosome 5 clone RP11-90P14, WORKING DRAFT SEQUENCE,

6 unordered pieces.

ACCESSION AC091992

VERSION AC091992.2 GI:15290436

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155276)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155276)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:14333928.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 441299
Center clone name: RPCI-II_90P14

Summary Statistics
Consensus quality: 151184 bases at least Q40
Consensus quality: 153297 bases at least Q30

COMMENT

Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23322236.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWTC
 Center clone name: CH230-436I19
 ----- Summary Statistics

Assembly program: Phrap; version 0.990329
 Consensus quality: 206076 bases at least Q40
 Consensus quality: 209800 bases at least Q30
 Consensus quality: 212212 bases at least Q20
 Estimated insert size: 201535; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      4140: contig of 4140 bp in length
*      4141: gap of unknown length
*      4241: contig of 10477 bp in length
*      14717: gap of unknown length
*      14817: gap of unknown length
*      14818: contig of 187852 bp in length
*      202670: gap of unknown length
*      202770: contig of 15316 bp in length
*      218085: gap of unknown length
*      218185: contig of 4832 bp in length
*      223017: gap of unknown length
*      223018: gap of unknown length
*      223118: contig of 1119 bp in length
*      224237: gap of unknown length
*      224337: contig of 1134 bp in length
*      225471: gap of unknown length
*      225571: contig of 2000 bp in length
*      227571: gap of unknown length
*      227671: contig of 1291 bp in length
*      228961: gap of unknown length
*      229061: contig of 1095 bp in length
*      229062: gap of unknown length
*      230157: contig of 1103 bp in length
*      230257: gap of unknown length
*      231360: contig of 1345 bp in length
*      231459: gap of unknown length
*      231460: contig of 1345 bp in length
*      232805: gap of unknown length
*      232804: contig of 1324 bp in length
*      232905: contig of 1324 bp in length

```

FEATURES
Source

```

1. .234228
   /organism="Rattus norvegicus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10116"
   /clone="CH230-436I19"
misc_feature 1. .1386
              /note="wgs_contig"

```

```

misc_feature 14818. .16195
              /note="wgs_contig"
misc_feature 200359. .202669
              /note="wgs_contig"
misc_feature 205939. .207188
              /note="wgs_contig"
misc_feature 214714. .215721
              /note="wgs_contig"
misc_feature 218186. .220168
              /note="wgs_contig"

ORIGIN
Query Match      9.1%; Score 42.4; DB 2; Length 234228;
Best Local Similarity 76.5%; Pred.No.3.7;
Matches 52; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10 CCTATCCACCTCCGCCAAGCCCTTTACCTATGCTGCTGCTAAACGCTGCTGCTGCTG 69
Db 86795 CTTCTCCCTCCCTCCCTCCCTCCCTCTTCTCTGCTGCTGCTGCTGCTGCTG 86736
QY 70 CTGCTGCT 77
Db 86735 CTGCTGCT 86728

```

Search completed: May 25, 2004, 16:30:27
 Job time : 3370 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 15:09:31 ; Search time 2857 Seconds
(without alignments)
4881.217 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED
Perfect score: 467
Sequence: 1 caatgttgcctaccct.....taataaacagttaaaagctg 457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estic:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| C 1 | 466 | 99.8 | 728 | 12 | BM975075 |
| C 2 | 462.2 | 99.0 | 555 | 10 | BE855577 |
| C 3 | 456 | 97.6 | 482 | 9 | AI871469 |
| C 4 | 454 | 97.2 | 517 | 14 | CA444588 |

```

78      47.2  10.1  1203  29  CNS015WU      AL106008 Drosophil
79      47.7  10.1  1299  10  AW485179   AL485179 Drosophil
80      46.8  10.0  1043  29  CNS0145P   AL103735 Drosophil
81      46.6  10.0  1201  13  BX374796   BX374796 Drosophil
82      46.2  9.9   394  13  BY086456   BY086456 Drosophil
83      46.2  9.9   693  13  BX381320   BX381320 Drosophil
84      46.2  9.9   997  29  CNS003AV   AL064073 Drosophil
85      46.6  9.9   843  29  CNS00CS1   AL059666 Drosophil
86      45.8  9.8   462  9   AA32984    LD12163.5
87      45.9  9.8   896  29  CNS008P8   AL057462 Drosophil
88      45.9  9.6  1201  29  CNS022AU   AL220719 Tetraodon
89      45.9  9.6  1201  9   AL513975   AL513975
90      44.6  9.6   672  28  BH396372   BH396372 AG-ND-159
91      44.4  9.5   844  29  CNS0052P   AL056652 Drosophil
92      44.2  9.5   660  28  BZ447727   BZ447727 BONOT32TR
93      43.8  9.4   509  14  CD342739   CD342739 EtESTee56
94      43.8  9.4   578  14  CD659850   CD659850 EtESTef37
95      43.8  9.4   778  29  CNS010UP   AL099451 Drosophil
96      43.8  9.4  1014  29  CNS0071P   AL0666705 Drosophil
97      43.8  9.4  1103  13  BX334917   BX334917
98      43.6  9.3   586  14  CD660017   CD660017 EtESTef38
99      43.2  9.3   381  13  BX360796   BX360796
100     43.2  9.3   728  29  CE142464   CE142464 tigr-gss-

```

ALIGNMENTS

```

RESULT 1
LOCUS BM975075/c
DEFINITION UI-CF-ECL-acf-1-23-0-UI.s1 UI-CF-ECL Homo sapiens cDNA clone
ACCESSION BM975075
VERSION BM975075.1 GI:19592566
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE discovery
PUBMED Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 408-444. >(CAG)n$Simple_repeat 685-728,
>(GGAA)n$Simple_repeat (matched complement)
Seq primer: M13_FORWARD
POLYA=Yes.
Location/Qualifiers
1..728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ECL-acf-1-23-0-UI"
/tissue_type="Lung"

```

FEATURES

```

source
1..728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ECL-acf-1-23-0-UI"
/tissue_type="Lung"

```

```

/dev_stage="Adult and Fetal"
/lab_host="PH108 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-ECL"
/notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-ECL is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGTCTTAC.
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_LIB=UI-CF-ECL
TAG_SEQ=AAGTGTAC"

```

ORIGIN

```

Query Match      99.8%; Score 466; DB 12; Length 728;
Best Local Similarity 100.0%; Pred. No. 3.6e-111;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CAATCTTTGGCTATCCACTCCCTCCCAAGCCCTTTACCTATGCTGCTGTAAACGCTGCTG 60
Db      484  CAATCTTTGGCTATCCACTCCCTCCCAAGCCCTTTACCTATGCTGCTGTAAACGCTGCTG 425

QY      61  CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGGACTGCTGGTCCGCCAGAAA 120
Db      424  CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGGACTGCTGGTCCGCCAGAAA 365

QY      121  GTCCTTTTGGCACTGACGCGCCCTCCATCAGGATGGGGCTTCTTTCCCTTCCTTTCTG 180
Db      364  GTCCTTTTGGCACTGACGCGCCCTCCATCAGGATGGGGCTTCTTTCCCTTCCTTTCTG 305

QY      181  TGCTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db      304  TGCTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245

QY      241  GAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGCTG 300
Db      244  GAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGCTG 185

QY      301  GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTATATCTTATATCTTATATCAA 360
Db      184  GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTATATCTTATATCTTATATCAA 125

QY      361  AGATATTAAAGCCCTGTTTCATTAAAGAAATCTCCCTTCCCTTCCCTGTTCAATGTTGTAA 420
Db      124  AGATATTAAAGCCCTGTTTCATTAAAGAAATCTCCCTTCCCTTCCCTGTTCAATGTTGTAA 65

QY      421  AGATTGTTCTGTGTAATATGCTTTTATAATATAACAGTTAAAGCT 466
Db      64  AGATTGTTCTGTGTAATATGCTTTTATAATATAACAGTTAAAGCT 19

```

```

RESULT 2
BE855577/c
LOCUS BE855577
DEFINITION 7g10e01.x1 NCI CGAP Brn23 Homo sapiens cDNA clone IMAGE:3306072 3,
similar to contains element OFR repetitive element ;, mRNA
sequence.
ACCESSION BE855577
VERSION BE855577.1 GI:10367753
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 555)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGA), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 472.
Location/Qualifiers
1. .555
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue="IMAGE:3306072"
/lab_host="glioblastoma (pooled)"
/clone_lib="NCI CGAP Brn23"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5',
TGTACCAATCTGAAGTGGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

FEATURES
source
Query Match 99.0%; Score 462.2; DB 10; Length 555;
Best Local Similarity 99.4%; Pred. No. 3.2e-110;
Matches 464; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATGTTTGGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTG 60
Db 468 CAATGTTTGGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTG 409
QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGTCGGTCCCAAGAA 120
Db 408 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGTCGGTCCCAAGAA 349
QY 121 GTCTCTTCTGCCACTGACGCCCCCATCAGGATGCGCTTCTTCCCTTCTTCTG 180
Db 348 GTCTCTTCTGCCACTGACGCCCCCATCAGGATGCGCTTCTTCCCTTCTTCTG 289
QY 181 TGTCTCTTCCCTCATCGGCTGCTATGCTGACCTGACGCAAGCCCGCTGGGAAGGG 240
Db 288 TGTCTCTTCCCTCATCGGCTGCTATGCTGACCTGACGCAAGCCCGCTGGGAAGGG 229
QY 241 GAGAAAGTGGGGATGGCTAAGAACTGGGAGATAGGGAACGAGAGGGTAGTGGGTG 300
Db 228 GAGAAAGTGGGGATGGCTAAGAACTGGGAGATAGGGAACGAGAGGGTAGTGGGTG 169
QY 301 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTACTAATTTATACAA 360
Db 168 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTACTAATTTATACAA 109
QY 361 AGATATTAAGGGCCCTGTTTCAATTAAGAAATTTGTTCCCTTCCCTTCTTCTG 420
Db 108 AGATATTAAGGGCCCTGTTTCAATTAATATTTGTTCCCTTCCCTTCTTCTG 49
QY 421 AGATTGTTCTGTGTAATATGCTTTTATATAACAGTTAAAGCTG 467

```

```

Db 48 AGATTGTTCTGTGTAATATGCTTTTATAATAACAGTTAAAGCTG 2

RESULT 3
AI871469/c
LOCUS
DEFINITION
AI871469 482 bp mRNA linear EST 07-MAR-2000
w167d09.x1 NCI CGAP Brn25 Homo sapiens cDNA clone IMAGE:2429969 3,
similar to contains element TAR1 repetitive element ;, mRNA
sequence.
ACCESSION
VERSION .AI871469.1 GI:5545518
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 482)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGA), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1409 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 450.
Location/Qualifiers
1. .482
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2429969"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn25"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5',
TGTACCAATCTGAAGTGGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

ORIGIN
Query Match 97.6%; Score 456; DB 9; Length 482;
Best Local Similarity 99.8%; Pred. No. 1.3e-108;
Matches 467; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CAATGTTTGGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTG 60
Db 468 CAATGTTTGGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTG 409
QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGTCGGTCCCAAGAA 120
Db 408 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGTCGGTCCCAAGAA 349
QY 121 GTCTCTTCTGCCACTGACGCCCCCATCAGGATGCGCTTCTTCCCTTCTTCTG 180
Db 348 GTCTCTTCTGCCACTGACGCCCCCATCAGGATGCGCTTCTTCCCTTCTTCTG 289

```

QY 181 TGCTCTCTGCTCATCGCCCTGCATGACCTGCAGCCCAAGCCAGCCCTGGGGAAGGG 240
 Db 288 TGCTCTCTGCTCATCGCCCTGCATGACCTGCAGCCCAAGCCAGCCCTGGGGAAGGG 229
 QY 241 GAG-AAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGGTAGTGGGT 299
 Db 228 GAGAAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGGTAGTGGGT 169
 QY 300 GGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATCTTATPACTAATTTATACA 359
 Db 168 GGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATCTTATPACTAATTTATACA 109
 QY 360 AAGATATTAAGGGCTGCTTATTAAGAAATGTTTCCCTTCCCTGGTGTCAATGTTTGTGA 419
 Db 108 AAGATATTAAGGGCTGCTTATTAAGAAATGTTTCCCTTCCCTGGTGTCAATGTTTGTGA 49
 QY 420 AAGATTTCTGTGTAAATATGCTTTATATAAAGAGTTTAAAGCTG 467
 Db 48 AAGATTGTTCTGTGTAATATGCTTTATATAAAGAGTTTAAAGCTG 1

RESULT 4
 CA444588/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 The following repetitive elements were found in this cDNA
 sequence: 409-448, >(CAG)n#simple_repeat
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1. .517
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-DTI-aw1-m-06-0-UI"
 /issue_type="Metastatic Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI-CGAP_DTI"

/notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI-CGAP_DTI is a normalized cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Ronaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this

library is AACTGTTCCG.
 TAG_TISSUE=lung metastatic chondrosarcoma
 TAG_LIB=UI-H-DTI
 TAG_SEQ=AACTGTTCCG"

ORIGIN

Query Match 97.2%; Score 454; DB 14; Length 517;
 Best Local Similarity 99.4%; Pred. No. 4.5e-108; Indels 3; Gaps 1;
 Matches 467; Conservative 0; Mismatches 0;
 QY 1 CAATGTTTCCCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGAAC---GCTG 57
 Db 488 CAATGTTTCCCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 429
 QY 58 CTGCTCTCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGTCGTCGTCGCCAG 117
 Db 428 CTGCTCTCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGTCGTCGTCGCCAG 369
 QY 118 AAGTCTCTTCTGCCACTGACGCCCCCATCAGGATTTGGGCTTCTTTCCCTCCCTCTTT 177
 Db 368 AAGTCTCTTCTGCCACTGACGCCCCCATCAGGATTTGGGCTTCTTTCCCTCCCTCTTT 309
 QY 178 CTGCTCTCTCTCTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237
 Db 308 CTGCTCTCTCTCTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
 QY 238 GGGGAGAAAGTGGGGTGGCTTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGG 297
 Db 248 GGGGAGAAAGTGGGGTGGCTTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGG 189
 QY 298 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATCTTATCTTATA 357
 Db 188 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATCTTATACTAAT 129
 QY 358 CAAAGATATTAAGGGCTGCTTATTAAGAAATGTTTCCCTTCCCTGCTGCTCAATGTTTG 417
 Db 128 CAAAGATATTAAGGGCTGCTTATTAAGAAATGTTTCCCTTCCCTGCTGCTCAATGTTTG 69
 QY 418 TAAAGATTTCTGTGTAAATATGCTTTTATAAAGAGTTTAAAGCTG 467
 Db 68 TAAAGATTTCTGTGTAAATATGCTTTTATAAAGAGTTTAAAGCTG 19

RESULT 5

BQ006616/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 The following repetitive elements were found in this cDNA
 sequence: 409-448, >(CAG)n#simple_repeat
 Seq primer: M13 FORWARD
 POLYA=Yes.

BQ006616.1 GI:19731516

Homo sapiens

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Jose Mercuende

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

The following repetitive elements were found in this cDNA

sequence: 409-448, >(CAG)n#simple_repeat

Seq primer: M13 FORWARD

POLYA=Yes.

Location/Qualifiers

Query Match 91.5%; Score 427.4; DB 13; Length 889;
 Best Local Similarity 97.8%; Pred. No. 5.1e-101;
 Matches 455; Conservative 0; Mismatches 6; Indels 4; Gaps 2;

QY 1 CAATGTTGGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTGCTGCTGCTGCTG 60
 DB 462 CAATGTTGGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTGCTGCTGCTGCTG 403

QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACCTGGTGGGTCGCCAGAAA 120
 DB 402 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACCTGGTGGGTCGCCAGAAA 343

QY 121 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
 DB 342 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 283

QY 181 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 DB 282 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 223

QY 241 GAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAAAGGCTGCTGCTG 300
 DB 222 GAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAAAGGCTGCTGCTG 163

QY 301 GGCTAGGGGGGCTGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360
 DB 162 GGCTAGGGGGGCTGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 103

QY 361 AGATATTAAAGGCTGCTGCTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 420
 DB 102 AGATATTAAAGGCTGCTGCTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 43

QY 421 AGATTTGTTCTGTGTAATATGCTTTATTAATAAAGCTGCTGCTGCTGCTGCTGCTGCT 465
 DB 42 AGA---TGTTGGTAAATGCTTTATTAATAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2

RESULT 9
 AI493364/c AI493364 437 bp mRNA linear EST 30-MAR-1999
 LOCUS tg70g07.x1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:2114172
 DEFINITION 3' similar to contains element MSRI repetitive element ;, mRNA sequence.

ACCESSION AI493364.1 GI:4394367
 VERSION AI493364
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 437)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 COMMENT Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov
 This clone is available royalty-free through LILNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 551. Std Error: 0.00
 Seq primer: -40UP from Gibco.
 Location/Qualifiers
 1. .437
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2114172"
 /tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
 /lab_host="DH10B"
 /clone_lib="Soares_NhMPu_S1"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I;

Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbM, pregnant uterus NhMPu, and fetal heart NhH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

ORIGIN

Query Match 90.5%; Score 422.8; DB 9; Length 437;
 Best Local Similarity 99.5%; Pred. No. 6.4e-100;
 Matches 424; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 42 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 101
 DB 437 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 378

QY 102 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 161
 DB 377 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318

QY 162 CTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 221
 DB 317 CTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 258

QY 222 CCAGCCCCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 281
 DB 257 CCAGCCCCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 198

QY 282 CAGAGAGGCTAGTGGTGGGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 341
 DB 197 CAAAAAGGCTAGTGGTGGGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 138

QY 342 CTTTACTAATTTATACAAAGATATTAGGCGCTGCTTCAATTAAGAAATTTGTTCCCTTCCC 401
 DB 137 CTTTACTAATTTATACAAAGATATTAGGCGCTGCTTCAATTAAGAAATTTGTTCCCTTCCC 78

QY 402 CTGTGTTCAATGTTTAAAGATTTGTTCTGTGTAATAATGTTTATATAAAGATTTAA 461
 DB 77 CTGTGTTCAATGTTTAAAGATTTGTTCTGTGTAATAATGTTTATATAAAGATTTAA 18

QY 462 AAGCTG 467
 DB 17 AAGCTG 12

RESULT 10
 AA099445
 LOCUS zk85f07.r1 Soares_pregnant uterus NhMPu Homo sapiens cDNA clone IMAGE:489637 5' similar to contains element PTRS repetitive element ;, mRNA sequence.

ACCESSION AA099445
 VERSION AA099445.1 GI:1645393
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 431)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 PUBMED 8889549

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 528 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham.
 Location/Qualifiers
 1. 431
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3804224"
 /db_xref="taxon:9606"
 /clone="IMAGE:489637"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="Soares_pregnant_uterus_NbHPU"
 /note="Organ: uterus; Vector: pVT73-Pac; Site 1: Not I;
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo (dT) primer [5',
 AACTGGAAGAAATCGCGCGCCCTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pVT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."

ORIGIN
 Query Match 90.0%; Score 420.2; DB 9; Length 431;
 Best Local Similarity 99.3%; Pred. No. 3.1e-99;
 Matches 422; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 7 | TTGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTG | 66 |
| Db | 6 | TTTGCTATCCACTCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTG | 65 |
| Qy | 67 | CTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGGTCGGTGCCTGCCAGAAATCTCT | 126 |
| Db | 66 | CTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGGTCGGTGCCTGCCAGAAATCTCT | 125 |
| Qy | 127 | TCTGCCACTGACGCCCCCATCAGGGATTGGCCCTCTTTCCCTCCCTCTCTTCTGTGCTC | 186 |
| Db | 126 | TCTGCCACTGACGCCCCCATCAGGGATTGGCCCTCTTTCCCTCCCTCTCTTCTGTGCTC | 185 |
| Qy | 187 | CTGCTCATGCGGCTGCCATGCACTGCGCCAGCCAGCCAGCCCTGGGGAGGGAGAA | 246 |
| Db | 186 | CTGCTCATGCGGCTGCCATGCACTGCGAGCCAGCCAGCCCTGGGGAGGGAGAA | 245 |
| Qy | 247 | GTGGGGATGCTAAGAAGCTGGGAGATAGGGACAGAGAGGGTAGTGGGTGGGCTAG | 306 |
| Db | 246 | GTGGGGATGCTAAGAAGCTGGGAGATAGGGACAGAGAGGGTAGTGGGTGGGCTAG | 305 |
| Qy | 307 | GGGGGCTGCCCTATTATAAGTGGTGTGTTATGATCTTATCTAATTTATACAAGATAT | 366 |
| Db | 306 | GGGGGCTGCCCTATTATAAGTGGTGTGTTATGATCTTATCTAATTTATACAAGATAT | 365 |
| Qy | 367 | TAAAGCCCTGTTCTATTAGAATTTGCCCTTCCCTGTTCAATGTTCTGAAGATTG | 426 |
| Db | 366 | TAAAGCCCTGTTCTATTAGAATTTGCCCTTCCCTGTTCAATGTTCTGAAGATTG | 425 |
| Qy | 427 | TTCTG 431 | |
| Db | 426 | TTCTG 430 | |

RESULT 11
 AA101878/c
 LOCUS
 DEFINITION zk85f07.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
 AA101878 458 bp mRNA linear EST 11-MAY-1997

IMAGE:489637 3', mRNA sequence.

AA101878 GI:1645281

EST.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 458)

AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Warra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

PUBMED 8889549

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 528 Std Error: 0.00
Seq primer: -40M13 fwd from Amersham.
Location/Qualifiers
1..458
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3804224"
/db_xref="taxon:9606"
/clone="IMAGE:489637"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares pregnant uterus NBHPU"
/note="Organ: uterus; Vector: pVT3-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5'-AATCGGAAGATTGGCGGCCTTTTTCCTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pVT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

FEATURES source

Query Match 89.8%; Score 419.2; DB 9; Length 458;
Best Local Similarity 97.3%; Pred.No. 5.7e-99;
Matches 435; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

Qy 22 CCCCAAGCCCTTTACCTATGCTGCTGTCTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 81

Db 458 CCCCAGCCCCCTTTACTATGCTTGCTTNNTAACGCTGCTGCTGCTGCTGCTGCTGCTTAAA 399

Qy 82 GGCT-CATGCTTGGAGTGGGAGCTGGTTCGGTCCCGCAGAAAGTCTCTTTGTGCCATGACGC 140

Db 398 GGCTCCAATGCTTGAGTGGGAGCTGGTTCGGTCCCGCAGAAAGTCTCTTTGTGCCATGACGC 339

Qy 141 CCCCATCAGGGATTGGGCTCTTTTCGCCCTTCCTTTCTGTGCTCTGCTGCTCATCGCC 200

Db 338 CCCCATCAGGGATTGGGCTCTTTTCGCCCTTCCTTTCTGTGCTCTGCTGCTCATCGCC 279

Qy 201 TGCCATGACCTGCAGCCAGCCAGCCCGCTGGGCGAAGGAGAGAAAGTGGGGGATGGCTA 260

Db 278 TGCCATGACCTGCAGCCAGCCAGCCCGCTGGGCGAAGGAGAGAAAGTGGGGGATGGCTA 219

Qy 261 AGAAGCTGGGAGATAGGGAACAAGAAGGGGTAGTGGGTGGGCTAGGGGGGTGCCATTAT 320

| | | | |
|----|-----|---|-----|
| Qy | 53 | CGTGTGCTGCTGCTGCTGCTGCTAAAGGCTCATGCTTGAGTGGGACATGTCGGTG | 112 |
| Db | 414 | CGCTNCTGCTGCTCCTGCTGCTNC-TAAAGGCTCATCTTGAAGTGGGAACATGGTTCGGTG | 356 |
| Qy | 113 | CCAGAGAAGTCTCTTCTGCCACTCAAGCCCCCATCAGGATTGGCCCTTCTTTCCCCCTT | 172 |
| Db | 355 | CUNANAAGTCTCTTCTGCCACTCAAGCCCCCATCAGGATTGGGCTTCTTTCCCCCTT | 296 |
| Qy | 173 | CTTTCTGTGTCTCCTGCTCATCGGCCTGCGCATGACCTGCAGCCCAAGCCCAGGCCCGCTG | 232 |
| Db | 295 | CTTTCTGTGTCTCCTGCTCATCGGCCTGCGATGACCTGCAGCCCAAGCCCAGGCCCGCTG | 236 |
| Qy | 233 | GGGAAGGGGAGAAAGTGTGGGATGGCTAAGAAAAGCTGGGAGATGGGAACAAGAGGGT | 292 |
| Db | 235 | GGGAAGGGGAGAAAGTGTGGGATGGCTAAGAAAAGCTGGGAGATGGGAACAAGAGGGT | 176 |
| Qy | 293 | AGTGGGTGGCTAGGGGGCTGCCTTAATTAAAGTGGTGTGTTATGATCTTATACTAAT | 352 |
| Db | 175 | AGTGGGTGGCTAGGGGGCTGCCTTAATTAAAGTGGTGTGTTATGANNCTATATACTAAT | 116 |
| Qy | 353 | TTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCTCTCCCTGTGTTCAAT | 412 |
| Db | 115 | TTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCTCTCCCTGTGTTCAAT | 56 |
| Qy | 413 | GTTTGTAAAGATGTCCTGCTAAATATGCTTTATATAAACACAGTTAAAAAGCTG | 467 |
| Db | 55 | GTTTGTAAAGATTCCTCTGTAAATATGCTTTATATAAACACAGTTAAAAAGCTG | 1 |

RESULT 19
AA083737/c
LOCUS
DEFINITION zn39d09.s1 Stracagene endothelial cell 937223 Homo sapiens CDNA clone IMAGE:549809 3', mRNA sequence.
ACCESSION AA083737
VERSION AA083737.1 GI:1625813
KEYWORDS EST.
SOURCE Homo sapiens (human)

| ORIGIN | Query Match | Best Local Similarity | Matches | Score | DB % | Length | 396; |
|--------|-------------|---|---------|-------|------|--------|------|
| | | | | | | | |
| QY | 77 | TTAAGGCTCATGTTGGAGTGGGACATGGTTCGGTGCCTCCAGAAAGTCTCTTCTGCCACTG | 336 | | | | |
| Db | 395 | TTAAAGGCTCATGCTTGGAGTGGGACATGGTTCGGTGCCTCCAGAAAGTCTCTTCTGCCACTG | 336 | | | | |
| QY | 137 | ACGCCCCCATCAGGGATTGGGCCCTCTTTTCCCTCTCTTTCTGTGTCTCTGTGCTCTATC | 196 | | | | |
| Db | 335 | ACGCCCCCATCAGGGATTGGGCCCTCTTTTCCCTCTCTTTCTGTGTCTCTGTGCTCTATC | 278 | | | | |
| QY | 197 | GGCTGTGCCATCCTGTCAGCCCAAGCCCGCCCTGGGGAGGGGAGAAAGTGGGGGATG | 256 | | | | |
| Db | 275 | GGCTGTGCCATGACCTGTCAGCCCAAGCCCGCCCTGGGGAGGGGAGAAAGTGGGGGATG | 216 | | | | |
| QY | 257 | GCTAAGAAAGCTGGAGATAGGGACAGAGAGGGTATGTGGTGGCTAGGGGGCTGCCC | 316 | | | | |
| Db | 215 | GCTAAGAAAGCTGGAGATAGGGACAGAGAGGGTATGTGGTGGCTAGGGGGCTGCCC | 156 | | | | |
| QY | 317 | TTATTTTAAAGTGGTTGTTTATGATTTCTATCTAAATTTATACAAAGATATTAAAGGCCCTG | 376 | | | | |
| Db | 155 | TTATTTTAAAGTGGTTGTTTATGATTTCTATCTAAATTTATACAAAGATATTAAAGGCCCTG | 96 | | | | |
| QY | 377 | TTCATTAAGAAATGTTTCCCTTCCCTCTGTTTCAATGTTTGTAAAGATGTTCTCTGTGTAA | 436 | | | | |
| Db | 95 | TTCATTAAGAAATGTTTCCCTTCCCTCTGTTTCAATGTTTGTAAAGATGTTCTCTGTGTAA | 36 | | | | |
| QY | 437 | ATATGCTTTTATATAAACAGTTAAAGCTG | 467 | | | | |
| Db | 35 | ATATGCTTTTATATAAACAGTTAAAGCTG | 5 | | | | |

```

REFERENCE
AUTHORS
1. (Bases 1 to 396)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Hillier, L., Lennon, G., Becker, M., Ronaldo, M.F., Chiapelli, B.,
Chisoso, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
and Marra, M.
TITLE
Generation and analysis of 280,000 human expressed sequence tags
JOURNAL
Genome Res. 6 (9), 807-828 (1996)
MEDLINE
97044478
COMMENT
8889549
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham.
Location/Qualifiers
1. .396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3928326"
/db_xref="taxon:9606"
/clone="IMAGE:549809"
/dev_stage="umbilical vein, 1 passage"
/lab host="SOIR (kanamycin resistant)"
/clone lib="Stratagene endothelial cell 937223"
FEATURES
source

```

RESULT 20
 AL570794/c
 LOCUS
 DEFINITION
 AL570794 Homo sapiens FLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 clone CSOD1022YC23 3-PRIME, mRNA sequence.
 ACCESSION
 AL570794
 VERSION
 AL570794.2 GI:31292209
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 977)
 Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 ON Feb 16, 2001 this sequence version replaced gi:12927452.
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: secre@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 10757.r For
 more information about this cluster, see
[http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CSOD1022AB12NP1&cluster=10757.r](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOD1022AB12NP1&cluster=10757.r). Contact :
 Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
 Paraday Avenue genoscope sequence ID : CSOD1022AB12NP1.
 Location/Qualifiers
 1..977
 /organism="Homo sapiens"
 source
 FEATURES

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YC23"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

```

Query Match      81.5%; Score 380.6; DB 9; Length 977;
Best Local Similarity 95.5%; Pred. No. 9.7e-89;
Matches 420; Conservative 3; Mismatches 13; Indels 4; Gaps 3;

QY 1 CAATGTTGGCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGCTA--ACGCTGC 58
Db 448 CAATGTTGGCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTTAACGCTGCTGC 389
QY 59 TGCCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGACTGCTGCTGCCAGA 118
Db 388 TGCCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGACTGCTGCTGCCAGA 329
QY 119 AAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178
Db 328 AAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
QY 179 TGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 238
Db 268 TGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 209
QY 239 GCGAGAAAGTGGGGATGGCTAGAAAGCTGGAGATAGGAGACAGAGAGGTAGTGG 298
Db 208 GCGAGAAAGTGGGGATGGCTAGAAAGCTGGAGATAGGAGACAGAGAGGTAGTGG 149
QY 299 TGGGCTAGGGGGCTGCTGCTTATTAAGTGGT--TGTTTATGATCTTATTAATTA 357
Db 148 TGGGCTAGGGGGCTGCTGCTTATTAAGTGGT--TGTTTATGATCTTATTAATTA 89
QY 358 CAAAGATATTAAAGCCCTGCTTAAAGAAATGTTCC--CCTCCCTGCTGCTCAATGTTT 416
Db 88 CAAAGATATTAAAGCCCTGCTTAAAGAAATGTTCC--CCTCCCTGCTGCTCAATGTTT 29
QY 417 GTAAGATTTGTTCTGTGTAA 436
Db 28 GTAAGATTTGTTCTGTAAA 9

```

```

RESULT 21
AW594286/c 386 bp mRNA linear EST 22-MAR-2000
LOCUS hg57h03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2949749 3',
DEFINITION mRNA sequence.
ACCESSION AW594286
VERSION AW594286.1 GI:7281544
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-i@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

```

```

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco.

```

FEATURES

```

Location/Qualifiers
1..386
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2949749"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

```

ORIGIN

```

Query Match      81.2%; Score 379.4; DB 10; Length 386;
Best Local Similarity 99.7%; Pred. No. 1.5e-88;
Matches 380; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 87 ATGCTTGGAGTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 146
Db 386 ATGCTTGGAGTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 327
QY 147 CAGGAGATGGGCTTCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 206
Db 326 CAGGAGATGGGCTTCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 267
QY 207 GACCTGCACCAAGCCAGCCGCTGGGAGAGGAGTGGGGATGGGCTAAGAAAG 266
Db 266 GACCTGCACCAAGCCAGCCGCTGGGAGAGGAGTGGGGATGGGCTAAGAAAG 207
QY 267 CTGGAGATAGGAGACAGAAAGAGGTAGTGGCTAGGGGGCTGCTTATTATAAG 326
Db 206 CTGGAGATAGGAGACAGAAAGAGGTAGTGGCTAGGGGGCTGCTTATTATAAG 147
QY 327 TGTGTTTATGATCTTATACATAATTATACAAAGATTAAGGCCCTGTTCAATTAAG 386
Db 146 TGTGTTTATGATCTTATACATAATTATACAAAGATTAAGGCCCTGTTCAATTAAG 87
QY 387 AATTGTTCCCTTCCCTGTTCAATGTTTGAAGATTTGTTCTGTGTAATATGCTTT 446
Db 86 AATTGTTCCCTTCCCTGTTCAATGTTTGAAGATTTGTTCTGTGTAATATGCTTT 27
QY 447 ATAATAAACAGTTAAAGCTG 467
Db 26 ATAATAAACAGTTAAAGCTG 6

```

RESULT 22

```

BQ024597/c 356 bp mRNA linear EST 27-MAR-2002
LOCUS UI-1-BB1p-aus-e-04-0-UI-s1 NCI_CGAP_P16 Homo sapiens cDNA clone
DEFINITION UI-1-BB1p-aus-e-04-0-UI 3', mRNA sequence.
ACCESSION BQ024597
VERSION BQ024597.1 GI:19759876
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1. .356
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UT-1-BB1p-aus-e-04-0-UI"
/tissue_type="Placenta"
/dev_stage="Full Term"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Pl6"
/note="Organ: Placenta; Vector: p77T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP Pl6 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p77T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are GA, AGGAA. For additional information, contact: Bento Soares, bento-soares@uiowa.edu
TAG_TISSUE=placenta human full term
TAG_LIB=UT-1-BB1p
TAG_SEQ=AGGAA"

ORIGIN
Query Match 72.4%; Score 338; DB 12; Length 356;
Best Local Similarity 100.0%; Pred. No. 1e-77;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 GCCACTGACGCCCATCAGGAGTGGCCCTCTTCCCTTCTCTTCTGTCCTCTG 189
DB 356 GCCACTGACGCCCATCAGGAGTGGCCCTCTTCCCTTCTCTTCTGTCCTCTG 297
QY 190 CCTCATCGCCCTGCATGACCTGCAGCCCAAGCCCGCCGCGGAGGAGGAAAGTG 249
DB 296 CTTATCGGCTGCATGACCTGCAGCCCAAGCCCGCCGCGGAGGAGGAAAGTG 237
QY 250 GGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGGGTAGGG 309
DB 236 GGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGGGTAGGG 177
QY 310 GGCTGCCTTATTTAAAGTGGTGTATGATCTTATCTATCTATCTATCTATCTAT 369
DB 176 GGCTGCCTTATTTAAAGTGGTGTATGATCTTATCTATCTATCTATCTATCTAT 117
QY 370 GGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGTTGTTCAATGTTTGAAGATTGTC 429
DB 116 GGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGTTGTTCAATGTTTGAAGATTGTC 57
QY 430 TGTGTAATATGCTCTTATTAATAAACAGTAAAGCTG 467
DB 56 TGTGTAATATGCTCTTATTAATAAACAGTAAAGCTG 19

RESULT 23
AI886944/c
LOCUS
DEFINITION w194d03.x1 NCI CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2432549 3',

mRNA sequence.
AI886944 GI:5592108
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (Bases 1 to 612)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 517 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 312.
Location/Qualifiers
1. .612
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2432549"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn25"
/note="Organ: brain; Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p77T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 65.0%; Score 303.6; DB 9; Length 612;
Best Local Similarity 93.1%; Pred. No. 1.3e-68;
Matches 363; Conservative 0; Mismatches 19; Indels 8; Gaps 4;
QY 78 TAAAGGCTCATGTGAGGTGGGAGCTGTCGTCGCCAGAAAGTCTCTCTGCACTGA 137
DB 384 TAAAGGCTCATGTGTT---GAGTGGGACTGTTCGTGCCAGAGGTCTCT--CTGCACTGA 330
QY 138 CGCCCCCATCAGGATTTGGCCCTTCTTCCCTTCTCTCTCTCTCTCTCTCTCTCT 197
DB 329 CCCCCCCCAACAGG--TTGGGCTTCTTCTTCCCTTCTCTCTCTCTCTCTCTCT 273
QY 198 GCCTGCCATCCTGCAGCCCAAGCCCGCTGGGGAAGGGGAGAAAGTGGGGATGG 257
DB 272 GCCTGCCATCCTGCAGCCCAAGCCCGCTGGGGAAGGGGAGAAAGTGGGGATGG 213
QY 258 CTAGAAAGCTGGAGATAGGGAACAGAAAGGTAGTGGGTAGGGCTAGGGGGCTGCCT 317
DB 212 CTAGAAAGCTGGAGATAGGGAACAGAAAGGTAGTGGGTAGGGGGCTGCCT 153
QY 318 TATTTAAAGTGGTCTTTATGATCTTATCTATCTATCTATCTATCTATCTATCTAT 377
DB 152 TATTTAAAGTGGTGTGTTATGATCTTATCTATCTATCTATCTATCTATCTATCTAT 93
QY 378 TCATTAAGAAATGTTCCCTTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 437


```

QY 377 TCCATTAAGAAATTCCTCCCTCCCTGGTTCCTCAATGTTGTAAGAGATTGTTCTGTGAA 436
Db |||||||
QY 91 TCCATTAAGAAATTCCTCCCTCCCTGGTTCCTCAATGTTGTAAGAGATTGTTCTGTGAA 32
Db |||||||
QY 437 ATATGCTCTTTATATAAACAAGTAAAGAGCTG 467
Db |||||||
31 ATATGCTCTTTATATAAACAAGTAAAGAGCTG 1

RESULT 26
AA304416 334 bp mRNA linear EST 18-APR-1997
LOCUS EST17255 Aorta endothelial cells, TNF alpha-treated Homo sapiens
DEFINITION CDNA 5' end, mRNA sequence.
ACCESSION AA304416
VERSION AA304416.1 GI:1956820
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 334)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.B., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirely,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Gao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen.H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
7566098
Other ESTs: THC174412
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@igr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
1..334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):116104"
/db_xref="taxon:9606"
/cell_type="endothelial cell"
/dev_stage="adult"
/clone_lib="Aorta endothelial cells, TNF alpha-treated"
/note="Organ: aorta; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
ORIGIN
Query Match 56.1%; Score 261.8; DB 9; Length 334;
Best Local Similarity 96.0%; Pred. No. 9.7e-58;
Matches 290; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

```

```

QY 1 CAATGTTGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTGTAACGCTGCTG 60
Db |||||||
QY 16 CAATGTTGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTGTAACGCTGCTG 75
Db |||||||
QY 61 CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAA 120
Db CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAA 135
QY 121 GTCTCTTCTGCACTGACGCGCCCATCAGGAGTTGGGCTCTTTCCTCCCTTCTCTG 180
Db |||||||
QY 136 GTCTCTTCTGCACTGACGCGCCCATCAGGAGTTGGGCTCTTTCCTCCCTTCTCTG 195
Db |||||||
QY 181 TGTCTCTGCTGCTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db |||||||
QY 196 TGTCTCTGCTGCTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255
QY 241 GAGAAAGT-GGGGGATGGCTAAGAAAGCT---GGGAGATAGGGACACAGAGAGGGTAGTG 296
Db |||||||
QY 256 GAGAAAGTGGGGGGATGGCTAAGAAAGCTTGGGAGATTAGGGGACCAAGAGGGTAGTT 315
Db |||||||
QY 297 GG 298
Db ||
316 GG 317

RESULT 27
BF939932/c
LOCUS BF939932
DEFINITION nacc64d01.x1 NCI CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3439129
3', mRNA sequence.
ACCESSION BF939932
VERSION BF939932.1 GI:12357252
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 300)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 292.
FEATURES
Location/Qualifiers
1..300
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3439129"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI-CGAP Brn23"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCACTGAAGTGGGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.

```

Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

```

ORIGIN
Query Match          53.7%; Score 250.8; DB 10; Length 300;
Best Local Similarity 93.5%; Pred. No. 7.1e-55;
Matches 261; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 189 GCTCATCGCCCTGCCATGACCTGAGCCAGCCAGCCCGCTGGGGAAGGGGAGAAAGT 248
DB 279 GCNCATGTTTCCCTATCCACCTCCCAAGCCCCCCCCCGTGGGGAAGGGGAGAAAGT 220

QY 249 GGGGATGGCTAAGAAAGCTGGGAGATAGGCAACAGAGAGGCTAGTGGGTGGCTAGGG 308
DB 219 GGGGATGGCTAAGAAAGCTGGGAGATAGGCAACAGAGAGGCTAGTGGGTGGCTAGGG 160

QY 309 GGGCTGCTTATTAAAGTGGTGTATATGATTCTTATATCTTATATCTTATATCAAAAGATATTA 368
DB 159 GGGCTGCTTATTAAAGTGGTGTATATGATTCTTATATCTTATATCTTATATCAAAAGATATTA 100

QY 369 AGGCCCTGTTCAATAGAAATGTTCCCTCCCTGTTCAATGTTTCTTAAAGATTGTT 428
DB 99 AGGCCCTGTTCAATAGAAATGTTCCCTCCCTGTTCAATGTTTCTTAAAGATTGTT 40

QY 429 CTGTGTAATATGCTCTTTATATAAAGAGTAAAGCTG 467
DB 39 CTGTGTAATATGCTCTTTATATAAAGAGTAAAGCTG 1

RESULT 28
AI031859/c
LOCUS
DEFINITION
OY45f12.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:166815 3',
mRNA sequence.
ACCESSION
AI031859
VERSION
AI031859.1 GI:3250071
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 409)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGAP), Tumor Gene Index
JOURNAL
Unpublished (1998)
AUTHORS
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 921 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 379.
FEATURES
Location/Qualifiers
1..409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:166815"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn23"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'

```

```

TGTACCAATCTGAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTTTTTTTT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

ORIGIN
Query Match          53.6%; Score 250.2; DB 9; Length 409;
Best Local Similarity 93.5%; Pred. No. 1.1e-54;
Matches 261; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 189 GCTCATCGCCCTGCCATGACCTGAGCCAGCCAGCCCGCTGGGGAAGGGGAGAAAGT 248
DB 284 GCCCAATGTTTGGCTATCCACTCCCCCAAGCCCCCGTGGGGAAGGGGAGAAAGT 225

QY 249 GGGGATGGCTAAGAAAGCTGGGAGATAGGCAACAGAGAGGCTAGTGGGTGGCTAGGG 308
DB 224 GGGGATGGCTAAGAAAGCTGGGAGATAGGCAACAGAGAGGCTAGTGGGTGGCTAGGG 165

QY 309 GGGCTGCTTATTAAAGTGGTGTATATGATTCTTATATCTTATATCAAAAGATATTA 368
DB 164 GGGCTGCTTATTAAAGTGGTGTATATGATTCTTATATCTTATATCAAAAGATATTA 105

QY 369 AGGCCCTGTTCAATAGAAATGTTCCCTCCCTGTTCAATGTTTCTTAAAGATTGTT 428
DB 104 AGGCCCTGTTCAATAGAAATGTTCCCTCCCTGTTCAATGTTTCTTAAAGATTGTT 45

QY 429 CTGTGTAATATGCTCTTTATATAAAGAGTAAAGCTG 467
DB 44 CTGTGTAATATGCTCTTTATATAAAGAGTAAAGCTG 6

RESULT 29
BF589966/c
LOCUS
DEFINITION
naa33h02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3258458
3', mRNA sequence.
ACCESSION
BF589966
VERSION
BF589966.1 GI:11682290
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 229)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
AUTHORS
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES
Location/Qualifiers
1..229
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3258458"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid11"
/notes="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP

```


TITLE Gene Expression Analysis in Response to Lung Toxicants: I.
 JOURNAL Sequencing and Microarray Development
 COMMENT Unpublished (2003)
 Contact: Shultz MA
 Dept. of Molecular Biosciences, School of Veterinary Medicine
 University of California, Davis
 1311 Haring Hall, One Shields Avenue, Davis, CA 95616, USA
 Tel: 530 752 0793
 Fax: 530 752 4698
 Email: mashultz@ucdavis.edu
 Average Phred score is 20 or better. All poor quality data (Phred < 20) and vector/linker sequence has been removed.
 High quality sequence stop: 799.

FEATURES Location/Qualifiers
 source 1..799
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="Contig2278"
 /sex="male"
 /tissue_type="airway or parenchyma"
 /dev_stage="adult"
 /clone_lib="Rat lung airway and parenchyma cDNA libraries"
 /note="Organ: lung; Vector: pGEM-11zf(-); Site 1: Eco RI; Site 2: Not I; mRNA was isolated from microdissected rat lung airways and parenchyma tissues."
 ORIGIN

Query Match 30.5%; Score 142.6; DB 14; Length 799;
 Best Local Similarity 73.8%; Pred. No. 2.3e-26;
 Matches 222; Conservative 0; Mismatches 74; Indels 5; Gaps 3;
 Qy 172 TCCTTTCTGTCTCCTCCCTCATCGGCTCCCATGACCTGCAGCCAGCCAGCCCGGT 231
 Db 8 TTCTCTGTCTTGTGTTTGTGTTTGTACCTCTTGCCTTATCAGCCTGCCATGGCAATCCAC 67
 Qy 232 GGGGAAGGGA--GAAGTGGGGATGCTAAGAAAGCTGGGAGATAGGACAGAGAG 289
 Db 68 AGGGAAGGGGAGGAGGAGGATGTGCTGAGAAAGAGGAGAGATAGAAACAGAGAGG 127
 Qy 290 GGTAGTGGTGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATCTTACT 349
 Db 128 GGGAGTGAATGGACCCAGTGGGCTGTCTTATTAAAGTGGTGTGTTATGATCTTACT 187
 Qy 350 AATTATACAAAGATATTAAGCCCTGTCTATT--AAGAAATGTTCCCTTCCCTGTGTT 408
 Db 188 AATTATATAGATATTAAGCCCTCTGAGTAAAGAAACTGTCTCATCCCGTGTGTT 247
 Qy 409 CA--ATGTTTGAAGATTGTTCTGTGTAATATGCTTTTATATAACAGTTAAAGCT 466
 Db 248 CACTATGTTTGAATAATGTTCCATGTAATATGCTTTTATATAAGAGTTAAAGCT 307
 Qy 467 G 467
 Db 308 G 308

RESULT 34
 BF599804
 LOCUS 263710 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 DEFINITION BF599804
 ACCESSION BF599804.1 GI:11696653
 VERSION
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 354)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrrenkrug,S.C.,

Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
 Chitko-McKown,C.G., Perte,G., Holt,I., Karamycheva,S., Liang,F.,
 Quackenbush,J. and Keefe,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 11282978
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCACGACG
 Plate: 35 row: P column: 11
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers

FEATURES source 1..354
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 3BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."
 ORIGIN

Query Match 27.8%; Score 130; DB 10; Length 354;
 Best Local Similarity 71.2%; Pred. No. 3.6e-23;
 Matches 259; Conservative 0; Mismatches 65; Indels 40; Gaps 5;
 Qy 110 GTGCCCAAGAAAGTCTCTTCTGCCACTGACGCCCCCATCAGGGATTGGGCTTCTTCCCT 169
 Db 2 GTGTCCAGAAAGCTCTTCTGCCATGGATGCCGCCACCAGGACTGGGCGTTTCCCT 61
 Qy 170 CTTCCTTCTGTGTCTCTCTGCTCATCGGCTGCCATGACCTGCGAGCCAGCCAGCCCT 229
 Db 62 CTTCCTCTCTCTCTCTCTCTGTCGATCAGTT-----GCCAGCCCT 102
 Qy 230 GTGGGAAGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAG 289
 Db 103 GTGGGAA--CGAGATGGAAGGAGTGGCTGGGAAAGTCTGGGACACGAGAGT-- 155
 Qy 290 GGTAGTGGTGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTACT 349
 Db 156 -----GGGGATGCTGAGGGGCTGTCTTATTTAAAGTGGTGTGTTAGTCTTAGACT 208
 Qy 350 AATTATACAAAGATATTAAGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTTCCCT 403
 Db 209 AATTATACAAAGTATTAAGGGCTTTTTCATTAAGAAAGCTTCCCTTTCCCTAGTTGT 268
 Qy 404 GTGTTCAATCTTGTGAAGATGTTCTGTGTAATATGCTTTATATAACAGTTAAAA 463
 Db 269 GTTCACCTGTGTTTGAAGATGTTCTAGTGTAAATATGTC--TTATATAAAGATGAAA 327
 Qy 464 GCTG 467
 Db 328 GCTG 331

RESULT 35
 BI901469
 LOCUS 520 bp mRNA linear EST 12-MAR-2002
 DEFINITION id20d03.y1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus

State Univ., from 2 l; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Patima Bonaldo.

ORIGIN

```

Query Match      27.8%; Score 129.8; DB 13; Length 520;
Best Local Similarity 69.0%; Pred. No. 4.6e-23;
Matches 245; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

Qy 121 GTCTCTTCTGCACTGACGCCGCCATCAGGATGGGCTTCTTCCCCCT-TCCTTCT 179
Db 123 GCCTCTTCTACTGCTGGTACCTACCAGGACCGGCTTCTCCGTTTGTGCTTCT 182
Qy 180 GTGTCTCTGCTCATCGCCCTGCGCATGACCTGAGCCAGCCAGCCCGCTGGGGAAG 239
Db 183 GCACCTCTGCTTATCAGCTGCGATGCGCCAGCCGCCACAGGAAAGGGGGGAATA 242
Qy 240 GGAGAAAGTGGGGATGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGT 299
Db 243 T-----TGCTGAGAAAGAGGAGATAGA-AGCACAGAGGGGAGTAAAC 287
Qy 300 GGGCTAGGGGGCTGCTTATTAAAGTGGTGTATTGATCTTATTAATAATTATACA 359
Db 288 GGGCCAGTGGGCTGCTTATTAAAGTGGTATTGATTTCTTATTAATAATTATATA 347
Qy 360 AAGATATTAAAGCCCTGTTCAATTAAGAAATTTGTTCCCTTCCC-----TGTGTCAAT 412
Db 348 GAGATATTAAAGCCCTTTGAGTAGAGAAAGTGTCCCTTCCCATAAGTGTGCTAT 407
Qy 413 GTTTGTAAGATTGTTCTGTGTAATATGCTTTATTAATAACAGTTAAAGCTG 467
Db 408 GTTTGTAAGAAATTTGTCATGTAATATGCTTTATTAATAAGAGTTAAAGTTG 462

```

```

RESULT 37
BI967057
LOCUS      521 bp      mRNA      linear      EST 12-MAR-2002
DEFINITION id27h04.y1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus
            musculus cDNA clone IMAGE:5664607 5', mRNA sequence.
ACCESSION  BI967057
VERSION     BI967057.1 GI:16341462
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS    Melton,D., Brown,J., Kenty,G., Permurt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
            Endocrine Pancreas Consortium
            Unpublished (2000)

```

```

TITLE      Endocrine Pancreas Consortium
JOURNAL     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
COMMENT     Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            MGI:1950933 This sequence now available from the IMAGE consortium,
            for clone orders contact: info@image.llnl.gov
            High quality sequence stop: 429.

```

FEATURES

```

source
1..521
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5664607"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH10B"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
N1-MMS1"
/notes="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an EcoT of 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library."

```

ORIGIN

```

Query Match      27.8%; Score 129.8; DB 12; Length 521;
Best Local Similarity 69.0%; Pred. No. 4.6e-23;
Matches 245; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

Qy 121 GTCTCTTCTGCACTGACGCCGCCATCAGGATGGGCTTCTTCCCCCT-TCCTTCT 179
Db 148 GCCTCTTCTACTGCTGGTACCTACCAGGACCGGCTTCTCCGTTTGTGCTTCT 207
Qy 180 GTGTCTCTGCTCATCGGCTGCGCATGACCTGAGCCAGCCAGCCCGCTGGGGAAG 239
Db 208 GCACCTCTGCTTATCAGCTGCGATGCGCCAGCCGCCACAGGAAAGGGGGGAATA 267
Qy 240 GGAGAAAGTGGGGATGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGT 299
Db 268 T-----TGCTGAGAAAGAGGAGATAGA-AGCACAGAGGGGAGTAAAC 312
Qy 300 GGGCTAGGGGGCTGCTTATTAAAGTGGTGTATTGATCTTATTAATAATTATACA 359
Db 313 GGGCCAGTGGGCTGCTTATTAAAGTGGTATTGATTTCTTATTAATAATTATATA 372
Qy 360 AAGATATTAAAGCCCTGTTCAATTAAGAAATTTGTTCCCTTCCC-----TGTGTCAAT 412
Db 373 GAGATATTAAAGCCCTTTGAGTAGAGAAAGTGTCCCTTCCCATAAGTGTGCTAT 432
Qy 413 GTTTGTAAGATTGTTCTGTGTAATATGCTTTATTAATAACAGTTAAAGCTG 467
Db 433 GTTTGTAAGAAATTTGTTCCATGTAATATGCTTTATTAATAAGAGTTAAAGTTG 487

```

```

RESULT 38
CF582778
LOCUS      966 bp      mRNA      linear      EST 24-SEP-2003
DEFINITION AGNCOURT 11363244 updated NIH MGC_137 Mus musculus cDNA clone
            IMAGE:6432441 5', mRNA sequence.
ACCESSION  CF582778
VERSION     CF582778.1 GI:35196040
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```


SOURCE ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 596)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 630 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 394.

FEATURES source
1..596
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1268316"
/db_xref="taxon:9606"
/clone="IMAGE:342941"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBH119W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGCGGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH119W."

ORIGIN
Query Match 27.2%; Score 127; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 2.6e-22; Mismatches 0; Indels 0; Gaps 0;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 341 TCTTATCTAATTATACAAAGATATTAAAGCCCTGTTCATTAAAGAAATGTTCCCTTCC 400
Db 1 TCTTATCTAATTATACAAAGATATTAAAGCCCTGTTCATTAAAGAAATGTTCCCTTCC 60
Qy 401 CCTGTGTCATGTTGTGAAGATGTTCTGTGTAATAATGCTTTTATAATAACAGTTA 460
Db 61 CCTGTGTCATGTTGTGAAGATGTTCTGTGTAATAATGCTTTTATAATAACAGTTA 120
Qy 461 AAAGCTG 467
Db 121 AAAGCTG 127

RESULT 44
LOCUS W84234
DEFINITION mf36f06.rl Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:407171 5', mRNA sequence.
ACCESSION W84234
VERSION W84234.1 GI:1541218
KEYWORDS EST.

SOURCE ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 486)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395365.
Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:250939
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 484.

FEATURES source
1..486
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:407171"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGCGGCATCTTTTCTTTTCTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match 27.1%; Score 126.6; DB 14; Length 486;
Best Local Similarity 68.5%; Pred. No. 3.1e-22; Mismatches 243; Conservative 0; Mismatches 89; Indels 23; Gaps 4;
Qy 121 GTCTCTTCTGCCATGACGCCCATCGGATTTGGGCTTCTTCCCTTCTTCTTCT 179
Db 121 GCCTCTTCTACTGTGTGTACCCCTTACCAGGACCGGGCTTCTCCGTTTGTGTTCT 180
Qy 180 GTGTCTCTCTCATCGGCTTGCCTGACCTGACGCAAGCCAGCCAGCCCGTGGGGAAGG 239
Db 181 GCACCTCTCGCTTATCAGCTGCCATGGCCGCCACAGGGAAGGGGGGAATA 240
Qy 240 GGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACGAGACAGGTAGTGGGT 299
Db 241 T-----TGGCTGAGAAAGAGAGAGATAGA-AGCACAAGAGAGGGGATAAC 285
Qy 300 GGGCTAGGGGGCTGCCCTATTATTAAGTGGTGTGTTATGATCTTATATAATTATACA 359
Db 286 GGGCCCCAGTGGCGTGTCTTATTAAAGTGGTATGTGTTATTTCTTATATAATTATATA 345
Qy 360 AAGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTC-----TGTGTTCAAT 412

```

346 GAGATATTAAGGCCCTTGAGTTAGAGAACTGTCCTCCCTCCCATAGTGTGACGCTAT 405
413 GTTTGTTAAAGATTCTTCGTGTAATATGCTTTTATATTAATAACAGTTAAAGCTG 467
|||||
406 GTTTGTTAAAGATTCTTCATGTAATATGCTTTTATATTAATAAGAGTTAAAGATTG 460

B8819942 449 bp mRNA linear EST 19-NOV-2001
B8819942 RIKEN full-length enriched, mammary gland RCB-0526
JYG-MC(A) cDNA Mus musculus cDNA clone G830003B15 3', mRNA
sequence.
B8819942
B8819942.1 GI:16992571
Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 449)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nomasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,K., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shero-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1. 449
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G830003B15"
/tissue_type="mammary gland"
/cell_line="RCB-0526 Jyg-MC(A)"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0526 Jyg-MC(A) cDNA"

FEATURES
source
1. 449

ORIGIN
source
1. 342

```

```

Query Match 26.8%; Score 124.2; DB 10; Length 449;
Best Local Similarity 68.7%; Pred. No. 1.3e-21;
Matches 252; Conservative 0; Mismatches 93; Indels 22; Gaps 5;

QY 106 GTCGGTGGCCAGAAAGTCTCTCTGCGACATGACGCCCATCAGGGATTGGGCTTCTTT 165
|||||
Db 95 GTTGGAGCTGGCTGGCGCTCTTCTACTGCTGGTACCCCTACAGGACCGGGCTTCTCC 154
|||||
QY 166 CCCCTT-TCCTTTCTGTCTCTGCTCTCATCGGCTGCCCATGACCTGCAGCCAAAGCCCA 224
|||||
Db 155 GTCTTTGTCTGCTTCGACCTCTCTCTTATCATGACCTGCCATG-----GCCAGGCCCC 207
|||||
QY 225 GCCCGTGGGGAAGGGAGAAAGTGGGGGATGCTGAAGCTGGGAGATAGGACAG 284
|||||
Db 208 ACAGGGAAGGGAGGGGGGAATAT-----TGCTGAGAAAGAGAGAGATAGA-AGCAC 260
|||||
QY 285 AAGAGGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTTTATGATTCCTT 344
|||||
Db 261 AAGAGGGGAGTAAACGGGCCAGTGGGCTGCTTATTTAAAGTGGTTATGTTATTCCTT 320
|||||
QY 345 ATACTAATTTATACAAAGATATTAAAGCCCTGTTCATTAGAATAATGTTCCCTTCCCC-- 402
|||||
Db 321 ATACTAATTTATATAGAGATATTAAAGCCCTTTGAGTTAGAGAACTGTCCTCTTCCCAT 380
|||||
QY 403 -----TGTGTTCAATGTTTGTAAAGATTGTTCTGTGTAAATATGCTTTTATAATAACAG 457
|||||
Db 381 AAGTGTGTTGCGTATGTTTTTAAANAATCTTCCATGTAATAATGCTTTTATAATAAGAG 440
|||||
QY 458 TTAAGAAG 464
|||||
Db 441 TTAAGAAG 447

RESULT 46
BF600467 342 bp mRNA linear EST 25-APR-2001
LOCUS 265082 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION
ACCESSION BF600467
VERSION BF600467.1 GI:11697624
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (Bases 1 to 342)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Fertez,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCCAAGTCACGACG
Plate: 37 row: P column: 11
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
source
1. 342

```

| FEATURES | SOURCE |
|---|--------|
| 1. The first two columns are labeled "FEATURES" and "SOURCE". | |
| 2. The third column is labeled "FEATURES". | |
| 3. The fourth column is labeled "SOURCE". | |
| 4. The fifth column is labeled "FEATURES". | |
| 5. The sixth column is labeled "SOURCE". | |
| 6. The seventh column is labeled "FEATURES". | |
| 7. The eighth column is labeled "SOURCE". | |
| 8. The ninth column is labeled "FEATURES". | |
| 9. The tenth column is labeled "SOURCE". | |
| 10. The eleventh column is labeled "FEATURES". | |
| 11. The twelfth column is labeled "SOURCE". | |
| 12. The thirteenth column is labeled "FEATURES". | |
| 13. The fourteenth column is labeled "SOURCE". | |
| 14. The fifteenth column is labeled "FEATURES". | |
| 15. The sixteenth column is labeled "SOURCE". | |
| 16. The seventeenth column is labeled "FEATURES". | |
| 17. The eighteenth column is labeled "SOURCE". | |
| 18. The nineteenth column is labeled "FEATURES". | |
| 19. The twentieth column is labeled "SOURCE". | |
| 20. The twenty-first column is labeled "FEATURES". | |
| 21. The twenty-second column is labeled "SOURCE". | |
| 22. The twenty-third column is labeled "FEATURES". | |
| 23. The twenty-fourth column is labeled "SOURCE". | |
| 24. The twenty-fifth column is labeled "FEATURES". | |
| 25. The twenty-sixth column is labeled "SOURCE". | |
| 26. The twenty-seventh column is labeled "FEATURES". | |
| 27. The twenty-eighth column is labeled "SOURCE". | |
| 28. The twenty-ninth column is labeled "FEATURES". | |
| 29. The thirtieth column is labeled "SOURCE". | |
| 30. The thirty-first column is labeled "FEATURES". | |
| 31. The thirty-second column is labeled "SOURCE". | |
| 32. The thirty-third column is labeled "FEATURES". | |
| 33. The thirty-fourth column is labeled "SOURCE". | |
| 34. The thirty-fifth column is labeled "FEATURES". | |
| 35. The thirty-sixth column is labeled "SOURCE". | |
| 36. The thirty-seventh column is labeled "FEATURES". | |
| 37. The thirty-eighth column is labeled "SOURCE". | |
| 38. The thirty-ninth column is labeled "FEATURES". | |
| 39. The fortieth column is labeled "SOURCE". | |
| 40. The forty-first column is labeled "FEATURES". | |
| 41. The forty-second column is labeled "SOURCE". | |
| 42. The forty-third column is labeled "FEATURES". | |
| 43. The forty-fourth column is labeled "SOURCE". | |
| 44. The forty-fifth column is labeled "FEATURES". | |
| 45. The forty-sixth column is labeled "SOURCE". | |
| 46. The forty-seventh column is labeled "FEATURES". | |
| 47. The forty-eighth column is labeled "SOURCE". | |
| 48. The forty-ninth column is labeled "FEATURES". | |
| 49. The fiftieth column is labeled "SOURCE". | |
| 50. The fifty-first column is labeled "FEATURES". | |
| 51. The fifty-second column is labeled "SOURCE". | |
| 52. The fifty-third column is labeled "FEATURES". | |
| 53. The fifty-fourth column is labeled "SOURCE". | |
| 54. The fifty-fifth column is labeled "FEATURES". | |
| 55. The fifty-sixth column is labeled "SOURCE". | |
| 56. The fifty-seventh column is labeled "FEATURES". | |
| 57. The fifty-eighth column is labeled "SOURCE". | |
| 58. The fifty-ninth column is labeled "FEATURES". | |
| 59. The sixtieth column is labeled "SOURCE". | |
| 60. The sixty-first column is labeled "FEATURES". | |
| 61. The sixty-second column is labeled "SOURCE". | |
| 62. The sixty-third column is labeled "FEATURES". | |
| 63. The sixty-fourth column is labeled "SOURCE". | |
| 64. The sixty-fifth column is labeled "FEATURES". | |
| 65. The sixty-sixth column is labeled "SOURCE". | |
| 66. The sixty-seventh column is labeled "FEATURES". | |
| 67. The sixty-eighth column is labeled "SOURCE". | |
| 68. The sixty-ninth column is labeled "FEATURES". | |
| 69. The seventieth column is labeled "SOURCE". | |
| 70. The seventy-first column is labeled "FEATURES". | |
| 71. The seventy-second column is labeled "SOURCE". | |
| 72. The seventy-third column is labeled "FEATURES". | |
| 73. The seventy-fourth column is labeled "SOURCE". | |
| 74. The seventy-fifth column is labeled "FEATURES". | |
| 75. The seventy-sixth column is labeled "SOURCE". | |
| 76. The seventy-seventh column is labeled "FEATURES". | |
| 77. The seventy-eighth column is labeled "SOURCE". | |
| 78. The seventy-ninth column is labeled "FEATURES". | |
| 79. The eightieth column is labeled "SOURCE". | |
| 80. The eighty-first column is labeled "FEATURES". | |
| 81. The eighty-second column is labeled "SOURCE". | |
| 82. The eighty-third column is labeled "FEATURES". | |
| 83. The eighty-fourth column is labeled "SOURCE". | |
| 84. The eighty-fifth column is labeled "FEATURES". | |
| 85. The eighty-sixth column is labeled "SOURCE". | |
| 86. The eighty-seventh column is labeled "FEATURES". | |
| 87. The eighty-eighth column is labeled "SOURCE". | |
| 88. The eighty-ninth column is labeled "FEATURES". | |
| 89. The ninetieth column is labeled "SOURCE". | |
| 90. The ninety-first column is labeled "FEATURES". | |
| 91. The ninety-second column is labeled "SOURCE". | |
| 92. The ninety-third column is labeled "FEATURES". | |
| 93. The ninety-fourth column is labeled "SOURCE". | |
| 94. The ninety-fifth column is labeled "FEATURES". | |
| 95. The ninety-sixth column is labeled "SOURCE". | |
| 96. The ninety-seventh column is labeled "FEATURES". | |
| 97. The ninety-eighth column is labeled "SOURCE". | |
| 98. The ninety-ninth column is labeled "FEATURES". | |
| 99. The hundredth column is labeled "SOURCE". | |

```

/db_xref="GDB:1323684"
/db_xref="taxon:9606"
/clone="IMAGE:415214"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal_liver_spleen_NFLS_S1"
/notes="Organ: Liver and Spleen; Vector: p7T73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
This is a subcloned version of the original Soares fetal
liver spleen NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
NACTGGAAGAATTATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

| | Query Match | 25.3%; | Score 118; | DB 14; | Length 616; |
|----|-----------------------|--|------------------|-----------|-------------|
| | Best Local Similarity | 100.0%; | Pred. No. 6e-20; | Indels 0; | Gaps 0; |
| | Matches 118; | Conservative 0; | Mismatches 0; | | |
| Qy | 350 | AATTTATACAAGATATTAAAGCCCTGTGTTCAATTAAGAAATTTGTTCCCTCCCTCGTGTTC | 409 | | |
| Db | 1 | AATTTATACAAGATATTAAAGCCCTGTGTTCAATTAAGAAATTTGTTCCCTCCCTCGTGTTC | 60 | | |
| Qy | 410 | AATGTTTGTAAAGATTGTTCTCTGTGTAATATGCTTTATAATAAACAGTTAAAGCTG | 467 | | |
| Db | 61 | AATGTTTGTAAAGATTGTTCTCTGTGTAATATGCTTTATAATAAACAGTTAAAGCTG | 118 | | |

| | | | | | |
|------------|---|-------------|------|--------|-----------------|
| LOCUS | BF231395 | 480 bp | mRNA | linear | EST 27-MAR-2003 |
| DEFINITION | 253781 BAC 5BOV Bos taurus cDNA 5', mRNA sequence. | | | | |
| ACCESSION | BF231395 | | | | |
| VERSION | BF231395.1 | GI:11170359 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | Bos taurus (cow) | | | | |
| ORGANISM | Bos taurus | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae; Bovinae; Bos. | | | | |
| REFERENCE | 1 (bases 1 to 480) | | | | |
| AUTHORS | Sonstegard,T., Capuco,A.V., White,J., Van Tassel,C.P., Connor,E.E., Cho J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and Quackenbush,J. | | | | |
| TITLE | Analysis of bovine mammary gland EST and functional annotation of the Bos taurus gene index | | | | |
| JOURNAL | Mamm. Genome 13 (7), 373-379 (2002) | | | | |
| MEDLINE | 22135956 | | | | |
| PUBMED | 12140684 | | | | |

| COMMENT |
|--|
| <p>Contact: Sonstegard TS USDA, ARS, Beltsville Agricultural Research Center Bldg. 200 Rm 2A, Beltsville, MD 20705, USA Tel: 301 504 8416 Fax: 301 504 8414 Email: tads@lpsl.barc.usda.gov</p> <p>Single pass sequencing. Bases called and alt trimmed with phred v0.980304.e. Vector identified by cross match with the -minscore 18</p> |

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACAG

BACKWARD: GTTTCCTCAGTACGACG

Plate: 105 row: I column: 14

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1. .480

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/tissue type="pooled"

/lab host="DH10B"

/clone lib="BARC 5BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled mRNA isolated from mammary

tissues at eight physiological, developmental, and disease

states."

ORIGIN

Query Match 25.2%; Score 117.8; DB 10; Length 480;
Best Local Similarity 69.6%; Pred. No. 6.2e-20;
Matches 218; Conservative 0; Mismatches 62; Indels 33; Gaps 3;

QY 89 GCTTGGAGTGGGACTGGTGGTGGCCAGAGAAAGTCTCTTCTGACACTGACGCCGCCCATCA 148
Db 201 GCCCGAGTGGGGCTGACCAAGTGTCCAGAAAGCCTCTCTGCCATGGATGCCCCACCA 260
QY 149 GGAATTGGGCTCTTTCCGCCCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208
Db 261 GGGACTGGGCGTTTTCCT 310
QY 209 CCTGCACCAAGCCAGCCCGTGGGAGAGGGGAGAGTGGGGATGGCTAGAGAGCT 268
Db 311 -----TTGCCAGCCCTGTGGGAA--CGAGATGGAAGGGATGGCTGGGAAGTC 359
QY 269 GGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGGGGCTGCTCTATTATAAGTG 328
Db 360 TGG-----GGACAGAGAGTGGGGATGCTGAGGGGCTGCTATTATAAGTG 407
QY 329 GTTGTATTATCTTATATTAATTAACAAGATTAAGCCCTGTTTCATTAAGAAA 388
Db 408 GTTGTGTGATCTTATAGACTAATTATACAAAGTTATTAAAGCCCTTTTCATTAAGAAA 467
QY 389 TTCTTCCTTCCC 401
Db 468 GCCTTCCTTCCC 480

RESULT 49
BI455596 921 bp mRNA linear EST 21-AUG-2001
LOCUS 603174854F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5257860 5',
DEFINITION mRNA sequence.
ACCESSION BI455596
VERSION BI455596.1 GI:15246252
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapsb-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLAM11651 row: k column: 13
High quality sequence stop: 655.

FEATURES

source

1. .921
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5257860"
/tissue type="tumor, gross tissue"
/dev stage="7 months"
/lab host="DH10B"
/clone lib="NCI_CGAP Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN

Query Match 24.8%; Score 115.8; DB 12; Length 921;
Best Local Similarity 69.9%; Pred. No. 2.6e-19;
Matches 248; Conservative 0; Mismatches 87; Indels 20; Gaps 6;

QY 121 GTCTCTTCTGCCACTGACGCCGCCCATCAGGATGGGCTTCTTCCCTTCTTCTG 180
Db 113 GCCTCTTCTACTGCTGTATACCCCTACCGGGACCGGGCTTC-----TCCTGTCTT 163
QY 181 TGTCTCCTCTCATCGGCTGCCATGACTGCAGCCAAAGCCAGCCCGTGGGGAAGGG 240
Db 164 TGTCTGTTTCTGCA--CTCTCTGCTTATCAGCTGCGCCAGCCACAGGGAAGG 222
QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTG 300
Db 223 GAG--GGGGGAATATTGGCTGAGAAAGAGGAGAGATAGA--AGCACAGAGAGGGAGTAAACG 280
QY 301 GCTAGGGGGCTGCTTATTAAAGTGGTGTGTTATGATCTTATTAATTAATTAACAA 360
Db 281 GGCCCAAGTGGGCTGCTTATTAAAGTGGTGTGTTATGATCTTATTAATTAATTAATAG 340
QY 361 AGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTCCCT-----TGTTTCAATG 413
Db 341 AGATATTAAAGGCCCTTGGATTAGAGAACTGTCCTTCCCATAGTGTGCTGATG 400
QY 414 TTGTGAA--AGATTGTTCTGTGTAATATATCTTTATTAATAACAGTTAAAGCTG 467
Db 401 TTGTGAA--AGATTGTTCTGTGTAATATATCTTTATTAATAACAGTTAAAGCTG 455

RESULT 50
AV365977 231 bp mRNA linear EST 14-NOV-1999
LOCUS AV365977 RIKEN full-length enriched, 16 days embryo lung Mus
DEFINITION musculus cDNA clone 8430414M03 3', mRNA sequence.
ACCESSION AV365977
VERSION AV365977.1 GI:6413624
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 231)
AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,
Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Orawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,
Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tominaga, N.,
Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yanai, A.,
Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE JOURNAL COMMENT

RIKEN Mouse ESTs (Komno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kusunagi,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp/>) for
further details.

FEATURES source

Location/Qualifiers
1..231
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="8430414M03"
/sex="mixed"
/tissue_type="lung"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 16 days embryo
lung"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATTAATATCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified phluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"

ORIGIN

Query Match 23.0%; Score 107.2; DB 9; Length 231;
Best Local Similarity 74.5%; Pred. No. 2.9e-17;
Matches 164; Conservative 0; Mismatches 48; Indels 8; Gaps 2;
QY 255 TGGCTAAGAGAGCTGGAGATAGGACAGAGAGGAGTAGTGGGTGGCTAGGGGGGCTG 314
DB 13 TGTCTAAGAAAAGAGAGAGATA-GAAGCACAGAGGGGAGTAAGCCGCGCTG 71
QY 315 CCTATTATTAAGTGGTGTCTTATGATCTTATCTAATTTATACAAAGATATTAAAGGCC 374
DB 72 TCTTATTAAAGTGGTATGTGTTCTTACTAATTTATATAGAGATATTAAAGGCC 131
QY 375 TGTTCAATTAAGAAATGTTCCTCCCTCC-----TGTTCAATGTGTGTAAGATTGT 427

Db 132 TTTCAGTTAGAGAAACTGTCCCTTCCCATAGTGTTCGCTATGTTTGTAAAAATTCT 191
QY 428 TCTGTGTAATATGTCCTTTTATAATAAACAGTTAAAGCTG 467
DB 192 TCCATGTAAATATTTTATATAATAAAGAGTTAAAGTTG 231
Search completed: May 25, 2004, 17:18.12
Job time : 2869 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 17:18:40 ; Search time 2864 Seconds
(without alignments)
4869.287 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caaagtgttcctatccacct.....taataaacagttaaaagctg 467

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 10

Total number of hits satisfying chosen parameters: 13639910

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_eston:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 466 | 99.8 | 728 | 12 | BM975075 UI-CF-EC1 |
| C 2 | 447 | 95.7 | 450 | 9 | AI149981 qf38h10.x |
| C 3 | 414 | 88.7 | 449 | 9 | AW015212 UI-H-B10p |
| C 4 | 414 | 88.7 | 517 | 14 | CA444588 UI-H-DT1- |

| | | | | | |
|------|-----|------|------|----|------------|
| C 5 | 414 | 88.7 | 668 | 12 | BQ006616 |
| C 6 | 384 | 82.2 | 555 | 10 | BE855577 |
| C 7 | 373 | 79.9 | 889 | 13 | EX434223 |
| C 8 | 370 | 79.2 | 431 | 9 | AA099445 |
| C 9 | 363 | 77.7 | 687 | 14 | CA438256 |
| C 10 | 347 | 74.3 | 482 | 9 | AI871469 |
| C 11 | 342 | 73.2 | 386 | 10 | AW594286 |
| C 12 | 338 | 72.4 | 356 | 12 | BQ024597 |
| C 13 | 332 | 71.1 | 428 | 10 | BF22471 |
| C 14 | 324 | 69.4 | 437 | 9 | AI93364 |
| C 15 | 291 | 62.3 | 406 | 10 | BE463932 |
| C 16 | 285 | 61.0 | 612 | 9 | AI886944 |
| C 17 | 242 | 51.8 | 300 | 10 | BF939932 |
| C 18 | 242 | 51.8 | 334 | 9 | AA304416 |
| C 19 | 242 | 51.8 | 409 | 9 | AI031859 |
| C 20 | 241 | 51.6 | 396 | 9 | AA083737 |
| C 21 | 241 | 51.6 | 458 | 9 | AA101878 |
| C 22 | 238 | 51.0 | 403 | 9 | AU156068 |
| C 23 | 205 | 43.9 | 977 | 9 | AL570794 |
| C 24 | 201 | 43.0 | 271 | 9 | AA367347 |
| C 25 | 197 | 42.2 | 414 | 14 | T27131 |
| C 26 | 183 | 39.2 | 492 | 14 | R60026 |
| C 27 | 156 | 33.4 | 1200 | 9 | AL571908 |
| C 28 | 155 | 33.2 | 1201 | 9 | AL570843 |
| C 29 | 132 | 28.3 | 229 | 10 | BF589966 |
| C 30 | 127 | 27.2 | 528 | 13 | EX118770 |
| C 31 | 127 | 27.2 | 596 | 14 | W67823 |
| C 32 | 118 | 25.3 | 616 | 14 | W91937 |
| C 33 | 81 | 17.3 | 544 | 10 | BF114974 |
| C 34 | 80 | 17.1 | 885 | 14 | CF456943 |
| C 35 | 75 | 16.1 | 348 | 9 | AA083736 |
| C 36 | 61 | 13.1 | 557 | 9 | AI871137 |
| C 37 | 58 | 12.4 | 556 | 12 | BM994192 |
| C 38 | 52 | 11.1 | 640 | 12 | BM992527 |
| C 39 | 52 | 11.1 | 744 | 12 | BQ006379 |
| C 40 | 41 | 8.8 | 574 | 13 | BQ234959 |
| C 41 | 34 | 7.3 | 491 | 9 | AI676745 |
| C 42 | 30 | 6.4 | 972 | 29 | CNS05DCI |
| C 43 | 29 | 6.2 | 311 | 9 | AA363384 |
| C 44 | 29 | 6.2 | 405 | 9 | AA065331 |
| C 45 | 29 | 6.2 | 423 | 29 | LBAP096B01 |
| C 46 | 29 | 6.2 | 482 | 14 | T58423 |
| C 47 | 29 | 6.2 | 732 | 12 | BG328052 |
| C 48 | 29 | 6.2 | 855 | 12 | BG528108 |
| C 49 | 28 | 6.0 | 323 | 12 | BJ600203 |
| C 50 | 28 | 6.0 | 345 | 10 | AW509816 |
| C 51 | 28 | 6.0 | 534 | 13 | BY479938 |
| C 52 | 28 | 6.0 | 572 | 10 | BF266891 |
| C 53 | 28 | 6.0 | 666 | 12 | BJ580384 |
| C 54 | 27 | 5.8 | 283 | 28 | AZ057146 |
| C 55 | 27 | 5.8 | 350 | 14 | D21888 |
| C 56 | 27 | 5.8 | 351 | 28 | BH701978 |
| C 57 | 27 | 5.8 | 393 | 14 | CD343253 |
| C 58 | 27 | 5.8 | 393 | 14 | CD343259 |
| C 59 | 27 | 5.8 | 449 | 9 | AU062636 |
| C 60 | 27 | 5.8 | 462 | 9 | AA392984 |
| C 61 | 27 | 5.8 | 462 | 10 | AW746753 |
| C 62 | 27 | 5.8 | 475 | 10 | BE363676 |
| C 63 | 27 | 5.8 | 486 | 10 | AW678716 |
| C 64 | 27 | 5.8 | 487 | 10 | AW746812 |
| C 65 | 27 | 5.8 | 497 | 9 | AI520474 |
| C 66 | 27 | 5.8 | 508 | 9 | AA539308 |
| C 67 | 27 | 5.8 | 509 | 10 | BE122310 |
| C 68 | 27 | 5.8 | 515 | 10 | AW678672 |
| C 69 | 27 | 5.8 | 525 | 14 | CD814478 |
| C 70 | 27 | 5.8 | 541 | 14 | CD344181 |
| C 71 | 27 | 5.8 | 554 | 12 | BI241290 |
| C 72 | 27 | 5.8 | 560 | 28 | AZ778502 |
| C 73 | 27 | 5.8 | 564 | 9 | AA950806 |
| C 74 | 27 | 5.8 | 572 | 9 | AA950805 |
| C 75 | 27 | 5.8 | 572 | 10 | AW679994 |
| C 76 | 27 | 5.8 | 575 | 14 | CF416326 |
| C 77 | 27 | 5.8 | 592 | 10 | AW745375 |

```

c 78      27      5.8      627      10      AW678825
c 79      27      5.8      639      12      BI638657
c 80      27      5.8      708      14      CD825475
c 81      27      5.8      736      10      BE391757
c 82      27      5.8      818      28      CNS020KB
c 83      27      5.8      867      28      BH719045
c 84      27      5.8      893      29      CG963374
c 85      27      5.8      915      29      CNS030H2
c 86      27      5.8      916      29      CNS01Q7J
c 87      27      5.8      948      13      BU906591
c 88      27      5.8      1355      13      BU716281
c 89      26      5.6      207      10      BB023936
c 90      26      5.6      217      9      AJ284913
c 91      26      5.6      257      9      AI256810
c 92      26      5.6      279      14      CD49837
c 93      26      5.6      280      14      CD343152
c 94      26      5.6      284      10      BF479424
c 95      26      5.6      295      9      AV137522
c 96      26      5.6      301      9      AU255321
c 97      26      5.6      310      13      BY468581
c 98      26      5.6      316      14      CD343666
c 99      26      5.6      317      12      BJ475346
c 100     26      5.6      317      28      AZ006229

```

ALIGNMENTS

```

RESULT 1
BM975075/c
LOCUS
DEFINITION
BM975075 728 bp mRNA linear EST 20-FEB-2003
UI-CF-EC1-acf-1-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
UI-CF-EC1-acf-1-23-0-UI 3', mRNA sequence.
BM975075
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 728)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8989548
Contact: McCray, PE
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@iowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 408-444, >(CAG)n#Simple repeat 685-728,
>(GGAA)n#Simple repeat (matched complement)
Seq primer: M13_FORWARD
POLYA=yes.
FEATURES
source
1. . 728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-acf-1-23-0-UI"
/tissue_type="Lung"

```

```

/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EC1"
/note="Organ: Lung; Vector: pT7T3-pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGTGGCTTAC.
TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG LIB=UI-CF-EC1
TAG SEQ=AAGTGGCTTAC"

```

ORIGIN

```

Query Match      99.8%; Score 466; DB 12; Length 728;
Best Local Similarity 100.0%; Pred. No. 1.2e-221; Indels 0; Gaps 0;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CAATGTTTGCTATCCACCTCCGCCCAAGCCCTTTACCTATGCTGTGCTAAGCTCTG 60
DB      484 CAATGTTTGCTATCCACCTCCGCCCAAGCCCTTTACCTATGCTGTGCTAAGCTCTG 425

QY      61  CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGTCGGTGCCAGAAA 120
DB      424 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGTCGGTGCCAGAAA 365

QY      121 GTCTCTCTGCCACTGACGCCGCCCATCAGGATTTGGCCCTTTCTCCCTCTCTCTTCTG 180
DB      364 GTCTCTCTGCCACTGACGCCGCCCATCAGGATTTGGCCCTTTCTCCCTCTCTCTTCTG 305

QY      181 TGCTCTCTGCTCATCGGCTGCTGCTGACCTGACCTGAGCCCAAGCCGCCCTGGGGAAGG 240
DB      304 TGCTCTCTGCTCATCGGCTGCTGCTGACCTGAGCCCAAGCCGCCCTGGGGAAGG 245

QY      241 GAGAAAGTGGGGCTGCTAAGAAAGCTGGAGATAGGGAACAGAGAGAGTGTGGTG 300
DB      244 GAGAAAGTGGGGCTGCTAAGAAAGCTGGAGATAGGGAACAGAGAGAGTGTGGTG 185

QY      301 GGCTAGGGGGCTGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTTATACAA 360
DB      184 GGCTAGGGGGCTGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTTATACAA 125

QY      361 AGATATTAGGCCCTGTTTCATTAAGAAATTTGCCCTTCCCTGCTGCTCAATGTTGTAA 420
DB      124 AGATATTAGGCCCTGTTTCATTAAGAAATTTGCCCTTCCCTGCTGCTCAATGTTGTAA 65

QY      421 AGATTGCTGCTGCTAATGCTTTTATAATAACAGTTAAAGCT 466
DB      64 AGATTGCTGCTGCTAATGCTTTTATAATAACAGTTAAAGCT 19

```

RESULT 2

```

AI149981/c
LOCUS
DEFINITION
AI149981 450 bp mRNA linear EST 10-NOV-1998
qf38h10.x1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1752355
3' similar to contains element PTRS repetitive element i, mRNA
sequence.
AI149981
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
The following repetitive elements were found in this cDNA sequence: 409-448, >(CAG)nSimple_repeat
Seq primer: M13 FORWARD
POLYA=yes.

Location/Qualifiers
1..668
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5846209"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Eri1"
/note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP Eri1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACACCTTGAC.

TAG TISSUE=chondrosarcoma
TAG LIB=UI-H-E11
TAG_SEQ=ACACTTGAC"

JOURNAL COMMENT

RESULTS

LOCUS DEFINITION

BE855577
7910e01.x1 NCI CGAP Brn23 Homo sapiens cDNA clone IMAGE:3306072 3,
similar to contains element OFR repetitive element ; mRNA
sequence.
BE855577
EST.
BE855577.1 GI:10367753
Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTCAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 472.
Location/Qualifiers
1..555
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3306072"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn23"
/notes="Organ: brain; Vector: pT7T3-D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCCGCATATCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match
Best Local Similarity 82.2%; Score 384; DB 10; Length 555;
Matches 100.0%; Pred. No. 1.1e-180; Indels 0; Gaps 0;

QY 1 CAATGTGCTATCCACCTCCCCAAGCCCTTTACCTATCTCTGCTAACGCTGCTG 60
Db 468 CAAATGTGCTATCCACCTCCCCAAGCCCTTTACCTATCTCTGCTAACGCTGCTG 409

QY 61 CTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGAGCTGCTGCCAGAAA 120
Db 408 CTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGAGCTGCTGCCAGAAA 349

QY 121 GTCTCTTCTGCCACTGAGCGCCCATCAGGAGTTGGGCTCTTTCCTTCCTTCCTG 180
Db 348 GTCTCTTCTGCCACTGAGCGCCCATCAGGAGTTGGGCTCTTTCCTTCCTTCCTG 289

QY 181 TGTCTCTGCTCATCGGCTGCGATGACCTGAGCCAAGCCAGCCCGCTGGGAGGG 240

Query Match
Best Local Similarity 88.7%; Score 414; DB 12; Length 668;
Matches 100.0%; Pred. No. 1.2e-195; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGAGCTGCTGCTG 113
Db 432 GCTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGAGCTGCTGCTG 373

QY 114 CCAGAAAGTCTCTTCTGCACTGACGCCCCCATCAGGAGTTGGCCCTTCTTCCCCTTC 173
Db 372 CCAGAAAGTCTCTTCTGCACTGACGCCCCCATCAGGAGTTGGCCCTTCTTCCCCTTC 313

QY 174 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 233
Db 312 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253

QY 234 GGAGGGGAGAAAGTGGGGAGTGGCTAAGAAGCTGGAGATAGGAGACAGAGAGGTA 293
Db 252 GGAGGGGAGAAAGTGGGGAGTGGCTAAGAAGCTGGAGATAGGAGACAGAGAGGTA 193

QY 294 GTGGGTGGGCTAGGGGGCTGCTATTATAAGTGGTGGTATTGATTTCTTACTAAAT 353
Db 192 GTGGGTGGGCTAGGGGGCTGCTATTATAAGTGGTGGTATTGATTTCTTACTAAAT 133

QY 354 TATACAAAGATATTAAGGCCCTGTTTCATTAAGAAATTTGTTCCCTTCCCTGTTCAAATG 413
Db 132 TATACAAAGATATTAAGGCCCTGTTTCATTAAGAAATTTGTTCCCTTCCCTGTTCAAATG 73

QY 414 TTGTAAAGATTGTTCTGTAAATATGCTTTTATAATAAACAGTTAAAGCTG 467

| | | | | | | | | |
|--|--|---|-----------------------------|--|--|--|--|--|
| Db | 288 | TGTTCTCTGCTCATCGCCCTGCCCCATGACCTTGACGCCAAGCCCGAGCCCCCTGTGGGGAAGGG | 222 | | | | | |
| Qy | 241 | GAGAAAGTGGGGATGGCTTAAGAAGCTGGGAGATAGGGAACAGAGAAGGGTAGTGGGTG | 300 | | | | | |
| Db | 228 | GAGAAAGTGGGGATGGCTTAAGAAGCTGGGAGATAGGGAACAGAGAAGGGTAGTGGGTG | 169 | | | | | |
| Qy | 301 | GGCTAGGGGGCTGCGCTATTATTAAGTGGTGTGTTATGATTTCTTATCTAATTTATACAA | 360 | | | | | |
| Db | 168 | GGCTAGGGGGCTGCGCTATTATTAAGTGGTGTGTTATGATTTCTTATCTAATTTATACAA | 109 | | | | | |
| Qy | 361 | AGATATTAAAGGCCCTGTTCATTAA | 384 | | | | | |
| Db | 108 | AGATATTAAAGGCCCTGTTCATTAA | 85 | | | | | |
| RESULT 7 | | | | | | | | |
| BX434223/c | | | | | | | | |
| LOCUS | BX434223 | 889 bp | mRNA linear EST 15-MAY-2003 | | | | | |
| DEFINITION | BX434223 | Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YJ24 | | | | | | |
| 3-PRIME, mRNA sequence. | | | | | | | | |
| ACCESSION | BX434223 | | | | | | | |
| VERSION | BX434223.1 | GI:30777248 | | | | | | |
| KEYWORDS | EST. | | | | | | | |
| SOURCE | Homo sapiens | (human) | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | | | |
| AUTHORS | 1. (bases 1 to 889) | | | | | | | |
| TITLE | Li, W.B., Gruber, C., Jessee, J. and Polayes, D. | | | | | | | |
| JOURNAL | Full-length cDNA libraries and normalization | | | | | | | |
| COMMENT | Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 10757.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAK025CH09NM1&cluster=10757.r. Contact : Feng Liang Email: fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0BAK025CH09NM1. Location/Qualifiers 1. .889 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DE009YJ24" /tissue_type="PLACENTA" /clone_lib="Homo sapiens PLACENTA" /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized." | | | | | | | |
| FEATURES | | | | | | | | |
| source | | | | | | | | |
| Query Match | | | | | | | | |
| Best Local Similarity 79.9%; Score 373; DB 13; Length 889; | | | | | | | | |
| Matches 423; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | |
| Qy | 1 | CAATGTTTGCTATCCACCTCCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG | 60 | | | | | |
| Db | 462 | CAATGTTTGCTATCCACCTCCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG | 403 | | | | | |
| Qy | 61 | CTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGTGGTTCGCGTCCCAAGAA | 120 | | | | | |
| Db | 402 | CTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGTGGTTCGCGTCCCAAGAA | 343 | | | | | |
| Qy | 121 | GTCTCTCTGCACTGACGGCCCCATCAGGATGGGCTTCTTTCCCTTCCTTCTG | 180 | | | | | |
| Db | 342 | GTCTCTCTGCACTGACGGCCCCATCAGGATGGGCTTCTTTCCCTTCCTTCTG | 283 | | | | | |


```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1409 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 450.

FEATURES
Location/Qualifiers
1..482
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2429969"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn25"
/note="Organ: brain; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCCGACATAGTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN
source
Query Match 74.3%; Score 347; DB 9; Length 482;
Best Local Similarity 99.8%; Pred. No. 3.4e-162;
Matches 467; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 CAATGTTTGCTATCCACCTCCCGACCCCTTACCTATGCTGCTAACGCTGCTG 60
468 CAATGTTTGCTATCCACCTCCCGACCCCTTACCTATGCTGCTAACGCTGCTG 409
61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCAGAA 120
408 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCAGAA 349
121 GTCTCTTCTGACATGACGCCCCCATAGGAGTGGGCTTCTTCCCTCTCTCTCTG 180
348 GTCTCTTCTGACATGACGCCCCCATAGGAGTGGGCTTCTTCCCTCTCTCTCTG 289
181 TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
288 TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 229
241 GAG-AAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTGGGT 299
228 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTGGGT 169
300 GGGCTAGGGGGCTGCTTAAAGTGGTGGTATGATTTCTTAACTAATTTATACA 359
168 GGGCTAGGGGGCTGCTTAAAGTGGTGGTATGATTTCTTAACTAATTTATACA 109
360 AAGATATTAAGGCCCTGCTTAAAGAAATGTTCCCTTCCCTGCTTCAATGTTTGT 419
108 AAGATATTAAGGCCCTGCTTAAAGAAATGTTCCCTTCCCTGCTTCAATGTTTGT 49

ORIGIN
source
Query Match 73.2%; Score 342; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.1e-159;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

87 ATGCTTGGAGTGGGACTGCTCGGTGCCAGAAAGTCTCTTCCACCTGACGCCCCCAT 146
386 ATGCTTGGAGTGGGACTGCTCGGTGCCAGAAAGTCTCTTCCACCTGACGCCCCCAT 327
147 CAGGAGTGGGCTTCTTCCCTTCCCTTCTCTGCTCTCTGCTGCTGCTGCTGCTG 206
326 CAGGAGTGGGCTTCTTCCCTTCTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCT 267
207 GACCTCAGCCAGCCCGCGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 266
266 GACCTCAGCCAGCCCGCGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 207
267 CTGGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGGGCTGCTTATTAAAG 356
206 CTGGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGGGCTGCTTATTAAAG 147

```

```

QY 327 TGGTGTGTTATGATCTTATCTTAATAATTATACAAAGATATTAAAGCCCTGTTCAATTAAGA 386
Db 146 TGGTGTGTTATGATCTTATATACTAATTTATACAAAGATATTAAAGCCCTGTTCAATTAAGA 87

QY 387 AATTGTTCCCTTCCCTGTTGTTCAATGTTTGAAGATTGTT 428
Db 86 AATTGTTCCCTTCCCTGTTGTTCAATGTTTGAAGATTGTT 45

RESULT 12
BQ024597/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ024597 356 bp mRNA linear EST 27-MAR-2002
UI-1-BB1p-aus-e-04-0-UI-s1 NCI CGAP P16 Homo sapiens cDNA clone
UI-1-BB1p-aus-e-04-0-UI 3', mRNA sequence.
BQ024597
BQ024597.1 GI:19759876
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 356)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1..356
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-1-BB1p-aus-e-04-0-UI"
/tissue_type="Placenta"
/dev_stage="Full Term"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP P16"
/note="Organ: Placenta; Vector: pT7T3-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP P16 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGGAA. For
additional information, contact: Bento Soares,
bento-soares@uiowa.edu
TAG TISSUE=placenta human full term
TAG LIB=UI-1-BB1p
TAG_SEQ=AGGAA"

ORIGIN
Query Match 72.4%; Score 338; DB 12; Length 356;
Best Local Similarity 100.0%; Pred. No. 1e-157;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GCCACTGAGCCCCCATCAGGATGGCCCTCTTCCCTTCTCTTCTGTCCTG 189
Db 356 GCCACTGAGCCCCCATCAGGATGGCCCTCTTCTCTTCTGTCCTG 297

```

```

QY 190 CCTCATCGCCCTGCATGACCTGCAGCCCAAGCCCGTGGGAAGGGAAGTG 249
Db 236 CCTCATCGCCCTGCATGACCTGCAGCCCAAGCCCGTGGGAAGGGAAGTG 237

QY 250 GGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTAGGG 309
Db 236 GGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTAGGG 177

QY 310 GGCTGCCCTATTAAAGTGGTGTATTATGATCTTATCTTATCTTATCTTATCTTAA 369
Db 176 GGCTGCCCTATTAAAGTGGTGTATTATGATCTTATCTTATCTTATCTTATCTTAA 117

QY 370 GGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTGTTCAATGTTTGAAGATTGTT 429
Db 116 GGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTGTTCAATGTTTGAAGATTGTT 57

QY 430 TGTGTAAATATGTTCTTTATATAATAACAGTTAAAGCTG 467
Db 56 TGTGTAAATATGTTCTTTATATAATAACAGTTAAAGCTG 19

RESULT 13
BQ222471/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ222471 428 bp mRNA linear EST 09-NOV-2000
hr75g11.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:3134372 3',
similar to contains element OFR repetitive element , mRNA
sequence.
BQ222471
BQ222471.1 GI:11129648
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
High quality sequence stop: 416.

FEATURES
source
1..428
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3134372"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid11"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

ORIGIN
Query Match 71.1%; Score 332; DB 10; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.1e-154;

```



```

/clone_lib="NCI CGAP Kid11"
/Note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

```

ORIGIN

```

Query Match      62.3%; Score 291; DB 10; Length 406;
Best Local Similarity 99.7%; Pred. No. 3.3e-134;
Matches 341; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 126 TTCTGCCACTGACGCCCCCATCAGGATGGGCTTCTTCCCTTCTTCTGTGCT 185
    |||||
Db 345 TTCTGCCACTGACGCCCCCATCAGGATGGGCTTCTTCCCTTCTTCTGTGCT 286

QY 186 CTTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGAGAA 245
    |||||
Db 285 CTTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGAGAA 226

QY 246 AGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTA 305
    |||||
Db 225 AGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTA 166

QY 306 GGGGGCTGCGCTTATTAAAGTGGTGTATGATCTTATCTACTAATTTATACAAAGATA 365
    |||||
Db 165 GGGGGCTGCGCTTATTAAAGTGGTGTATGATCTTATCTACTAATTTATACAAAGATA 106

QY 366 TTAAGGCCCTGTTCATTAAGAAATGTTCCTTCCCTGTGTTCAATGTTGTAAAGATT 425
    |||||
Db 105 TTAAGGCCCTGTTCATTAAGAAATGTTCCTTCCCTGTGTTCAATGTTGTAAAGATT 46

QY 426 GTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 467
    |||||
Db 45 GTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 4

```

```

RESULT 16
AI886944/c
LOCUS      w194d03.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2432549 3',
DEFINITION mRNA sequence.
ACCESSION AI886944.1 GI:5592108
VERSION    AI886944.1
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 517 Std Error: 0.00
Seq primer: -40UP from Gibco

```

```

FEATURES
source      High quality sequence stop: 312.
            Location/Qualifiers
            1..612

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2432549"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Brn25"

```

```

/Note="Organ: Brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGGCATAGGTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

```

ORIGIN

```

Query Match      61.0%; Score 285; DB 9; Length 612;
Best Local Similarity 100.0%; Pred. No. 3.4e-131;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 TCTCTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGGA 242
    |||||
Db 287 TCTCTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGGA 228

QY 243 GAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGTGGG 302
    |||||
Db 227 GAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGTGGG 168

QY 303 CTAGGGGGCTGCGCTTATTAAAGTGGTGTGTTATGATCTTATCTACTAATTTATCAAAG 362
    |||||
Db 167 CTAGGGGGCTGCGCTTATTAAAGTGGTGTGTTATGATCTTATCTACTAATTTATCAAAG 108

QY 363 ATATTAGGCCCTGTTCATTAAGAAATGTTCCTTCCCTGTGTTCAATGTTGTAAAG 422
    |||||
Db 107 ATATTAGGCCCTGTTCATTAAGAAATGTTCCTTCCCTGTGTTCAATGTTGTAAAG 48

QY 423 ATTGTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 467
    |||||
Db 47 ATTGTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 3

```

```

RESULT 17
BF939932/c
LOCUS      nac64d01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3439129
DEFINITION 3', mRNA sequence.
ACCESSION BF939932.1 GI:12357252
VERSION    BF939932.1
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 517 Std Error: 0.00
Seq primer: -40UP from Gibco

```

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -400P from Gibco

High quality sequence stop: 292.

FEATURES

```

source
1..300
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:3439129"
  /tissue_type="gliblastoma (pooled)"
  /lab_host="DH10B"
  /clone_lib="NCL CGAP_Brn23"
  /note="Organ: Brain; Vector: pT7T3D-Pac (Pharmacia) with a
  modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
  strand cDNA was primed with a Not I - oligo(dT) primer [5',
  TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTATTTTATTTTATTTTATTTT
  T 3']; double-stranded cDNA was ligated to Eco RI
  adaptors (Pharmacia), digested with Not I and cloned into
  the Not I and Eco RI sites of the modified pT7T3 vector.
  Library is normalized, and was constructed by Bento
  Soares and M.Fatima Bonaldo."
```

ORIGIN

```

Query Match      51.8%; Score 242; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. No. 9.6e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 226 CCCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGA 285
Db 242 CCCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGA 183

Qy 286 AGAGGTAGTGGTGGGCTAGGGGGCGCGCTTATTTAAAGTGTTGTTATGATTCCTTA 345
Db 182 AGAGGTAGTGGTGGGCTAGGGGGCGCGCTTATTTAAAGTGTTGTTATGATTCCTTA 123

Qy 346 TACTAAATTTACAAAGATATTAAGCCCTGCTTATTAAAGAAATGTTCCCTTCCCTGT 405
Db 122 TACTAAATTTACAAAGATATTAAGCCCTGCTTATTAAAGAAATGTTCCCTTCCCTGT 63

Qy 406 GTTCAATGTTTGTAAAGATGTTCTGTGTAATATGCTTTATATAAACAAGTTAAAGC 465
Db 62 GTTCAATGTTTGTAAAGATGTTCTGTGTAATATGCTTTATATAAACAAGTTAAAGC 3

Qy 466 TG 467
Db 2 TG 1
```

RESULT 18

```

AA304416
LOCUS EST17255 Aorta endothelial cells, TNF alpha-treated Homo sapiens
DEFINITION cDNA 5' end, mRNA sequence.
ACCESSION AA304416
VERSION AA304416.1 GI:1956820
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

REFERENCE

```

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
```

Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.-I., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseitine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Nature 377 (6547 Suppl), 3-174 (1995)
7566098
Other ESTs: THC174412
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/cdb/hgi/hgi.html>)

Seq primer: M13 Reverse

Location/Qualifiers

source

```

1..334
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="ATCC (inhost):116104"
  /db_xref="taxon:9606"
  /cell_type="endothelial cell"
  /dev_stage="adult"
  /clone_lib="Aorta endothelial cells, TNF alpha-treated"
  /note="Organ: aorta; Vector: pBluescript SK-; Site 1:
  EcoRI; Site 2: XhoI"
```

ORIGIN

```

Query Match      51.8%; Score 242; DB 9; Length 334;
Best Local Similarity 100.0%; Pred. No. 9.7e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 TATCCACCTCCCAAGCCCTTACTATGCTGCTTAACGCTGCTGCTGCTGCTGCT 71
Db 27 TATCCACCTCCCAAGCCCTTACTATGCTGCTTAACGCTGCTGCTGCTGCTGCT 86

Qy 72 GCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGCTGGTGGCCAGAAAGTCTTCTG 131
Db 87 GCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGCTGGTGGCCAGAAAGTCTTCTG 146

Qy 132 CACTGAGCCCCCATCAGGATTTGGGCTTCTTTCCCTTCTTCTTCTTCTGCTGCC 191
Db 147 CACTGAGCCCCCATCAGGATTTGGGCTTCTTTCCCTTCTTCTTCTTCTGCTGCC 206

Qy 192 TCAATCGGCTTGCATGACCTGACCCCAAGCCAGCCCGCTGGGGAAGGGAGAAAGTGGG 251
Db 207 TCAATCGGCTTGCATGACCTGACCCCAAGCCAGCCCGCTGGGGAAGGGAGAAAGTGGG 266

Qy 252 GG 253
Db 267 GG 268
```

RESULT 19

```

AI031859/c
LOCUS oy45f12.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1668815 3',
DEFINITION mRNA sequence.
ACCESSION AI031859
VERSION AI031859.1 GI:3250071
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 409)
 AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail@nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/btgp/image/image.html
 Insert Length: 921 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 379.
 Location/Qualifiers
 1..409
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1668815"
 /tissue_type="glioblastoma (pooled)"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Brn23"
 /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo (dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGCATATCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
 ORIGIN
 Query Match 51.8%; Score 242; DB 9; Length 409;
 Best Local Similarity 100.0%; Pred. No. 1e-109;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 226 CCCGCTGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGA 285
 Db 247 CCCGCTGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGA 188
 Qy 286 AGAGGGTAGTGGGTGGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATTCCTTA 345
 Db 187 AGAGGGTAGTGGGTGGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATTCCTTA 128
 Qy 346 TACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTTCCCTTCCCTGT 405
 Db 127 TACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTTCCCTTCCCTGT 68
 Qy 406 GTTCAATGTTTGTAAAGATTTCTCTGTGTAATATGCTTTTATATAAAGCAGTTAAAGC 465
 Db 67 GTTCAATGTTTGTAAAGATTTCTCTGTGTAATATGCTTTTATATAAAGCAGTTAAAGC 8
 Qy 466 TG 467
 Db 7 TG 6
 RESULT 20
 AA083737 396 bp mRNA linear EST 21-OCT-1996
 LOCUS zn39d09.s1 StrataGene endothelial cell 937223 Homo sapiens cDNA
 DEFINITION clone IMAGE:549809 3', mRNA sequence.
 AA083737
 ACCESSION AA083737
 VERSION AA083737.1 GI:1625813
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 396)
 AUTHORS Hillier L., Lennon G., Becker M., Bonaldo M.F., Chiapelli B., Chissoe S., Dietrich N., DuBuque T., Pavello A., Gish W., Hawkins M., Hultman M., Kucaba T., Lacy M., Le M., Le N., Mardis E., Moore B., Morris M., Parsons J., Prange C., Rifkin L., Rohlfing T., Schellenberg K., Soares M.B., Tan F., Thierry-Mieg J., Trevaskis E., Underwood K., Wohlmann P., Waterston R., Wilson R. and Marra M.
 TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 PUBMED 8889549
 COMMENT Contact: Willson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40M13 fwd from Amersham.
 Location/Qualifiers
 1..396
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3928326"
 /db_xref="taxon:9606"
 /clone="IMAGE:549809"
 /dev_stage="umbilical vein, 1 passage"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene endothelial cell 937223"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Umbilical vein endothelial cells, passaged once. Average insert size: 1.0 Kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTWTTTTTTTTTTTT 3'"
 ORIGIN
 Query Match 51.6%; Score 241; DB 9; Length 396;
 Best Local Similarity 100.0%; Pred. No. 3.1e-109;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 227 CCCCTGGGGAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAA 286
 Db 245 CCCGTGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAA 186
 Qy 287 GAGGGTAGTGGGTGGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATTCCTAT 346
 Db 185 GAGGGTAGTGGGTGGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATTCCTAT 126
 Qy 347 ACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTTCCCTTCCCTGTG 406
 Db 125 ACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTTCCCTTCCCTGTG 66
 Qy 407 TTCAATGTTTGTAAAGATTTCTCTGTGTAATATGCTTTTATATAAAGCAGTTAAAAAGCT 466
 Db 65 TTCAATGTTTGTAAAGATTTCTCTGTGTAATATGCTTTTATATAAAGCAGTTAAAAAGCT 6
 Qy 467 G 467
 Db 5 G 5
 RESULT 21
 AA101878/c 458 bp mRNA linear EST 11-MAY-1997
 LOCUS zk85f07.s1 Soares pregnant uterus_NBHPU Homo sapiens cDNA clone
 DEFINITION IMAGE:489637 3', mRNA sequence.

| | | | |
|-----------------------|---|---|---------------------|
| ACCESSION | AA101878 | 1 | GI:1645281 |
| VERSION | AA101878.1 | 1 | GI:1645281 |
| KEYWORDS | EST. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | | | |
| AUTHORS | Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schallenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasaki,E., Underwood,K., Wohlmann,P., Waterston,K., Wilson,R. and Marra,M. | | |
| TITLE | Generation and analysis of 280,000 human expressed sequence tags | | |
| JOURNAL | Genome Res. 6 (9), 807-828 (1996) | | |
| MEDLINE | 97044478 | | |
| PUBMED | 8889549 | | |
| COMMENT | Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 528 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham. | | |
| FEATURES | | | |
| source | 1. 458 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GB:3804224" /db_xref="taxon:9606" /clone="IMAGE:489637" /sex="female" /dev_stage="adult" /lab_host="DH10B" /clone_lib="Soares_pregnant_uterus_NBHPU" /note="Organ: uterus; Vector: pTV73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo (dt) primer [5', AACTGGAGAATTGCGGCGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo." | | |
| ORIGIN | | | |
| Query Match | 51.6%; | Score 241; | DB 9; Length 458; |
| Best Local Similarity | 100.0%; | Pred. No. 3.2e-109; | |
| Matches 241; | Conservative | 0; Mismatches | 0; Indels 0; Gaps 0 |
| QY | 227 | CCCGTGGGAGAGGGAGAGAAAGTGGGGGATGCGTAAAGAAAGCTGGAGATAGGGAACAGAA | 286 |
| Db | 252 | CCCGTGGGAGAGGGAGAGAAAGTGGGGGATGCGTAAAGAAAGCTGGAGATAGGGAACAGAA | 193 |
| QY | 287 | GAGGGTAGTGGGTGGGCTAGGGGGGCTGCCCTATTATTAAGTGGTGTGTTATGATCTTAT | 346 |
| Db | 192 | GAGGGTAGTGGGTGGGCTAGGGGGGCTGCCCTATTATTAAGTGGTGTGTTATGATCTTAT | 133 |
| QY | 347 | ACTAATTTATACAAGATATTAAAGCCCTGTTCTATTAAAGAAATGTTCCCTCCCTCGTG | 406 |
| Db | 132 | ACTAATTTATACAAGATATTAAAGCCCTGTTCTATTAAAGAAATGTTCCCTCCCTCGTG | 73 |
| QY | 407 | TTCAATGTTTGAAGATGTTCTGTGTPAATATGTCCTTTATATATAACAGTTTAAAGCT | 466 |
| Db | 72 | TTCAATGTTTGAAGATGTTCTGTGTPAATATGTCCTTTATATATAACAGTTTAAAGCT | 13 |
| QY | 467 | G 467 | QY |

| | | | |
|--|---|--|--|
| AUTHORS TITLE JOURNAL COMMENT | Sikela, J.M. Human cDNAs from infant brain Unpublished (1993) Contact: Sikela JM Department of Pharmacology University of Colorado Health Sciences Center Box C236, 4200 E. 9th Ave, Denver CO 80262-0236 Tel: 3032708637 Fax: 3032707097 Email: nikki@tally.uchsc.edu Seq primer: -21m13 Universal. Location/Qualifiers | | |
| | FEATURES source | | |
| | 1. .414 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="L1AB311B03" /clone_lib="Infant brain, LLNL array of Dr. M. Soares INIB" /notes="Vector: lafmid BA; Site 1: HindIII; Site 2: NotI; Normalized infant brain cDNA library made by Dr. M. Soares (Columbia University), oligo-dt primed and directionally cloned between HindIII (5') and NotI (3') sites" | | |
| ORIGIN | Query Match 42.2%; Score 197; DB 14; Length 414; Best Local Similarity 99.1%; Pred. No. 3.1e-87; Matches 347; Conservative 0; Mismatches 3; Indels 0; Gaps 0; | | |
| | QY 118 AAGTCTCTTCTGCCACTGACGCCCCCATCAGGATTTGGSCCTTCTTCCCTCTCTTT 177 Db 350 AAGTCTCTTCTGCCACTGACGCCCCCATCAGGATTTGGSCCTTCTTCCCTCTCTTT 291 | | |
| | QY 178 CTGTGTCTCTGCTCATCGGCTGCGCATGACCTGACGCCAAGCCAGCCCGCTGGGGAA 237 Db 290 CTGTGTCTCTGCTCATCGGCTGCGCATGACCTGACGCCAAGCCCGCTGGGGAA 231 | | |
| QY | 238 GGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGAAACAGAGAGGTAGTGG 297 Db 230 GGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGAAACAGAGAGGTAGTGG 171 | | |
| | QY 298 GTGGGCTAGGGGGCTGCCCTATTAAAGTGGTTGTTATGATCTTTACTAATTATA 357 Db 170 GTGGGCTAGGGGGCTGCCCTATTAAAGTGGTTGTTATGATCTTTACTAATTATA 111 | | |
| | QY 358 CAAAGATATTAAAGCCCTGTTCATTAAAGAAATTTTCCTTCCCTTCCCTGTTCATCTTTG 417 Db 110 CAAAGATATTAAAGCCCTGTTCATTAAAGAAATTTTCCTTCCCTGTTCATCTTTG 51 | | |
| QY | 418 TAAAGATTGTTCTGTGTAATATGCTTTATATAAACAGTTAAAGCTG 467 Db 50 TAAAGATTGTTCTGTGTAATATGCTTTATATAAACAGTTAAAGCTG 1 | | |
| RESULT 26 R60026/c LOCUS | R60026 492 bp mRNA linear EST 24-MAY-1995 YH12C03.s1 Soares infant brain INIB Homo sapiens cDNA clone IMAGE:43014 3' similar to contains MSRL repetitive element ;, mRNA sequence. | | |
| | ACCESSION R60026 VERSION R60026.1 GI:830721 KEYWORDS EST. SOURCE Homo sapiens (human) ORGANISM Homo sapiens | | |
| | REFERENCE 1 (bases 1 to 492) AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucab, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R. | | |
| TITLE JOURNAL COMMENT | The WashU-Merck EST Project Unpublished (1995) Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 355 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: SP6 High quality sequence stop: 363. Location/Qualifiers | | |
| | FEATURES source | | |
| | 1. .492 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:415555" /db_xref="taxon:9606" /clone="IMAGE:43014" /sex="female" /dev_stage="73 days post natal" /lab_host="DH10B (ampicillin resistant)" /clone_lib="Soares infant brain INIB" /note="Organ: whole brain; Vector: Lafmid BA; Site 1: Not I; Site 2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AAGTGAAGATTCGCGCGCCGAGGAATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo." | | |
| ORIGIN | Query Match 39.2%; Score 183; DB 14; Length 492; Best Local Similarity 99.6%; Pred. No. 3.1e-80; Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | |
| | QY 227 CCCGTGGGAAAGGAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAA 286 Db 252 CCCGTGGGAAAGGAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAA 193 | | |
| | QY 287 GAGGGTAGTGGGTGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATCTTAT 346 Db 192 GAGGGTAGTGGGTGGCTAGGGGGCTGCCTATTATTAAGTGGTGTGTTATGATCTTAT 133 | | |
| QY | 347 ACTAATTTATACAAGATATTAAAGCCCGCTTCATTAAAGAAATGTTCCCTTCCCTGTG 406 Db 132 ACTAATTTATACAAGATATTAAAGCCCGCTTCATTAAAGAAATGTTCCCTTCCCTGTG 73 | | |
| | QY 407 TTCAATGTTTGTAAAGATTGTTCTGTGTAAATATGCTTTTATAATAACAGTTA 460 Db 72 TTCAATGTTTGTAAAGATTGTTCTGTGTAAATATGCTTTTATAATAACAGTTA 19 | | |
| RESULT 27 AL571908/c LOCUS | AL571908 1200 bp mRNA linear EST 31-MAY-2003 clone CS0D1029YL01 3-PRIME, mRNA sequence. | | |
| | ACCESSION AL571908 VERSION AL571908.2 GI:31293299 KEYWORDS EST. SOURCE Homo sapiens (human) ORGANISM Homo sapiens | | |
| | REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1200) AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished (2001) JOURNAL | | |


```

/clone="IMAGE:342941"
/sex="unknown"
/dev stage="19 weeks"
/lab host="DH10B (ampicillin resistant)"
/clone lib="Soares fetal heart NBH19W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCCGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."

ORIGIN
Query Match      27.2%; Score 127; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 3.1e-52;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 TCTTATCTACTAATTATACAAAGATATTAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCC 400
Db 1 TCTTATCTACTAATTATACAAAGATATTAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCC 60

QY 401 CCTGTGTTCAATGTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATAATAAACAGTTA 460
Db 61 CCTGTGTTCAATGTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATAATAAACAGTTA 120

QY 461 AAAGCTG 467
Db 121 AAAGCTG 127

RESULT 32
W91937
LOCUS zh47c08.r1 Soares fetal liver spleen INFLS_S1 Homo sapiens cDNA
DEFINITION clone IMAGE:415214 5', mRNA sequence.
ACCESSION W91937
VERSION W91937.1 GI:1424298
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 616)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Weg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
MEDLINE 8889549
PUBMED
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estowatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 720 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 451.
FEATURES
Location/Qualifiers
1..616
source

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1323684"
/db_xref="taxon:9606"
/clone="IMAGE:415214"
/sex="male"
/dev stage="20 week-post conception fetus"
/lab host="DH10B (ampicillin resistant)"
/clone lib="Soares fetal liver spleen INFLS_S1"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
This is a subcloned version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AATCGGAAGAAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match      25.3%; Score 118; DB 14; Length 616;
Best Local Similarity 100.0%; Pred. No. 9.8e-48;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 AATTATACAAAGATATTAAAGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGTGTC 409
Db 1 AATTATACAAAGATATTAAAGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGTGTC 60

QY 410 AATGTTTCTGAAGATGTTCTGTGTAATAATGTTCTTTATAATAAACAGTTAAAGCTG 467
Db 61 AATGTTTCTGAAGATGTTCTGTGTAATAATGTTCTTTATAATAAACAGTTAAAGCTG 118

RESULT 33
BF114974/c
LOCUS hr72a04.x1 NCI-CGAP_Kid1l Homo sapiens cDNA clone IMAGE:313998 3',
DEFINITION mRNA sequence.
ACCESSION BF114974
VERSION BF114974.1 GI:10984450
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 544)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 468.
FEATURES
Location/Qualifiers
1..544
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:313998"
/lab host="DH10B"
/clone lib="NCI CGAP Kid1l"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

```

a modified polylinker; Site 1: Not I; Site 2: Eco RI; plasmid DNA from the normalized library NCI_CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo.

ORIGIN

Query Match 17.3%; Score 81; DB 10; Length 544;
Best Local Similarity 100.0%; Pred. No. 3e-29;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 387 AATGTTCCCTCCCTGTTCAATGTTGTAAGATTGTTCTGTTAAATATGCTTT 446
Db 544 AATGTTCCCTCCCTGTTCAATGTTGTAAGATTGTTCTGTTAAATATGCTTT 485
QY 447 ATAATAACAGTTAAAGCTG 467
Db 484 ATAATAACAGTTAAAGCTG 464

RESULT 34
CF456943
LOCUS 885 bp mRNA linear EST 04-SEP-2003
DEFINITION AGENCOURT 15337517 Human Anterior Horn Homo sapiens cDNA clone
IMAGE:30514432 5', mRNA sequence.
ACCESSION CF456943
VERSION CF456943.1 GI:34456599
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM577 row: 0 column: 17
High quality sequence start: 4
High quality sequence stop: 598.

FEATURES

source
1. .885
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:30514432"
/tissue_type="Peripheral Nervous system"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="Human Anterior Horn"
/note="Vector: pCMV-SPORT5.1; Site_1: EcoRV (destroyed); Site_2: NotI; Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb. Library was constructed by Invitrogen."

ORIGIN

Query Match 17.1%; Score 80; DB 14; Length 885;
Best Local Similarity 100.0%; Pred. No. 1e-28;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTAAACGCTG 60
Db 624 CAATGTTGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTAAACGCTG 683
QY 61 CTGCTGCTGCTGCTGCTTAA 80
Db 684 CTGCTGCTGCTGCTGCTTAA 703

RESULT 35
AA083736
LOCUS 348 bp mRNA linear EST 21-OCT-1996
DEFINITION zN39409.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA
clone IMAGE:549809 5', mRNA sequence.
ACCESSION AA083736
VERSION AA083736.1 GI:1625812
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 348)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissee, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R., and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
889549

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 324.

FEATURES

source

1. .348
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3928326"
/db_xref="taxon:9606"
/clones="IMAGE:549809"
/dev_stage="umbilical vein, 1 passage"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene endothelial cell 937223"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site_2: XbaI; Cloned unidirectionally. Primer: Oligo dT.
Umbilical vein endothelial cells, passaged once. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGGCAG 3' ~3' adaptor sequence: 5'
CTCGAGTTTATTTTTTTTTT 3'"

ORIGIN

Query Match 16.1%; Score 75; DB 9; Length 348;
Best Local Similarity 100.0%; Pred. No. 2.8e-26;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 TTCTTTCCCTTCCTTTCTGTTCTTCCTGCTCAFCGCGCTGCCATGACCTGACGCCAA 219
Db 147 TTCTTTCCCTTCCTTTCTGTTCTTCCTGCTCAFCGCGCTGCCATGACCTGACGCCAA 206
QY 220 GCCAGCCCGCTGGG 234
|||||

| | |
|------------------------|---|
| | |
| <hr/> | |
| 207 GCCACGCCCGTGGG 221 | |
| <hr/> | |
| Db | |
| RESULT 36 | |
| AI871137/c | |
| LOCUS | AI871137 557 bp mRNA linear EST 07-MAR-2000 |
| DEFINITION | wf79e09.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2431144 3', mRNA sequence. |
| ACCESSION | AI871137 |
| VERSION | AI871137 |
| KEYWORDS | EST. |
| SOURCE | AI871137.1 GI:5545105 |
| ORGANISM | Homo sapiens (human) |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. |
| TITLE | National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index |
| JOURNAL | Unpublished (1998) |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.llnl.gov/bbrp/image/image.html Insert length: 669 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 455. |
| FEATURES | location/Qualifiers |
| source | 1..557 |
| | /organism="Homo sapiens" |
| | /mol_type="mRNA" |
| | /db_xref="taxon:9606" |
| | /clone="IMAGE:2431144" |
| | /tissue_type="anaplastic oligodendroglioma" |
| | /lab_host="DH10B" |
| | /clone_lib="NCI CGAP Brn25" |
| | /note="Organ: brain; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTCGGCGCCGCATAGGTCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo." |
| ORIGIN | |
| Query Match | 13.1%; Score 61; DB 9; Length 557; |
| Best Local Similarity | 100.0%; Pred.No. 2.9e-19; |
| Matches | 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| QY | 375 TGTTCAATTAAGAANTGTTCCTCCCTGCTTCATGTTTGAAGATTGTTCTGTG 434 |
| Db | 557 TGTTCAATTAAGAANTGTTCCTCCCTGCTTCATGTTTGAAGATTGTTCTGTG 498 |
| QY | 435 A 435 |
| Db | 497 A 497 |
| RESULT 37 | |
| BM994192/c | |
| LOCUS | BM994192 556 bp mRNA linear EST 17-JUN-2002 |
| DEFINITION | UI-H-D70-atx-n-09-0-UI.s1 NCI CGAP_DTO Homo sapiens cDNA clone IMAGE:5865800 3', mRNA sequence. |
| ACCESSION | BM994192 |
| VERSION | BM994192 |
| KEYWORDS | EST. |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index |
| JOURNAL | Unpublished (1997) |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: Dr. Jose Mercuende cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Seg primer: M3 FORWARD POLYA-Yes. |
| FEATURES | location/Qualifiers |
| source | 1..556 |
| | /organism="Homo sapiens" |
| | /mol_type="mRNA" |
| | /db_xref="taxon:9606" |
| | /clone="IMAGE:5865800" |
| | /tissue_type="Metastatic Chondrosarcoma" |
| | /dev_stage="Adult" |
| | /lab_host="DH10B (Life Technologies)" |
| | /clone_lib="NCI CGAP DTO" |
| | /note="Organ: Lung; Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP DTO is a cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTV73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTCGCG. |
| | TAG_TISSUE=lung metastatic chondrosarcoma |
| | TAG_LIB=UI-H-DTO |
| | TAG_SEQ=AACTGTCGCG |
| ORIGIN | |
| Query Match | 12.4%; Score 58; DB 12; Length 556; |
| Best Local Similarity | 100.0%; Pred.No. 9.3e-18; |
| Matches | 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| QY | 410 AATGTTTGTAAAGATTGTTCTGTGTAATAATGCTTTAATAAACAGTTAAAGCTG 467 |
| Db | 537 AATGTTTGTAAAGATTGTTCTGTGTAATAATGCTTTAATAAACAGTTAAAGCTG 480 |
| RESULT 38 | |
| BM992527/c | |
| LOCUS | BM992527 640 bp mRNA linear EST 17-JUN-2002 |
| DEFINITION | UI-H-D70-att-k-01-0-UI.s1 NCI_CGAP_DTO Homo sapiens cDNA clone IMAGE:5864184 3', mRNA sequence. |
| ACCESSION | BM992527 |
| VERSION | BM992527 |
| KEYWORDS | EST. |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index |
| JOURNAL | Unpublished (1997) |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: Dr. Jose Mercuende cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Seg primer: M3 FORWARD POLYA-Yes. |
| FEATURES | location/Qualifiers |
| source | 1..556 |
| | /organism="Homo sapiens" |
| | /mol_type="mRNA" |
| | /db_xref="taxon:9606" |
| | /clone="IMAGE:5865800" |
| | /tissue_type="Metastatic Chondrosarcoma" |
| | /dev_stage="Adult" |
| | /lab_host="DH10B (Life Technologies)" |
| | /clone_lib="NCI CGAP DTO" |
| | /note="Organ: Lung; Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP DTO is a cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTV73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTCGCG. |
| | TAG_TISSUE=lung metastatic chondrosarcoma |
| | TAG_LIB=UI-H-DTO |
| | TAG_SEQ=AACTGTCGCG |
| ORIGIN | |
| Query Match | 12.4%; Score 58; DB 12; Length 556; |
| Best Local Similarity | 100.0%; Pred.No. 9.3e-18; |
| Matches | 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| QY | 410 AATGTTTGTAAAGATTGTTCTGTGTAATAATGCTTTAATAAACAGTTAAAGCTG 467 |
| Db | 537 AATGTTTGTAAAGATTGTTCTGTGTAATAATGCTTTAATAAACAGTTAAAGCTG 480 |
| RESULT 39 | |
| BM994192/c | |
| LOCUS | BM994192 556 bp mRNA linear EST 17-JUN-2002 |
| DEFINITION | UI-H-D70-atx-n-09-0-UI.s1 NCI CGAP_DTO Homo sapiens cDNA clone IMAGE:5865800 3', mRNA sequence. |
| ACCESSION | |

AUTHORS
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 The following repetitive elements were found in this cDNA sequence: 1-21, >AT rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes

FEATURES
 source
 1..640
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5864184"
 /tissue_type="Metastatic Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP DT0"
 /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Ecor I; Site_2: Not I; NCI CGAP DT0 is a cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACGTGTCGG.
 TAG TISSUE=lung metastatic chondrosarcoma
 TAG LIB=UI-H-DT0
 TAG_SEQ=AACTGTCGG"

ORIGIN
 Query Match 11.1%; Score 52; DB 12; Length 640;
 Best Local Similarity 100.0%; Pred. No. 9.4e-15;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TGTAAAGATTGTTCTGTTAAATATGCTTTTATATAACAGTTTAAAGCTG 467
 |||||
 Db 617 TGTAAAGATTGTTCTGTTAAATATGCTTTTATATAACAGTTTAAAGCTG 566
 |||||

RESULT 39
 BQ006379/c
 LOCUS
 DEFINITION UI-H-E11-aza-i-11-0-UI.s1 NCI CGAP_E11 Homo sapiens cDNA clone
 IMAGE:5846038 3', mRNA sequence.
 ACCESSION BQ006379
 VERSION BQ006379.1 GI:19731279
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homidae; Homo.
 1 (bases 1 to 744)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

AUTHORS
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 The following repetitive elements were found in this cDNA sequence: 1-21, >AT rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes

FEATURES
 source
 1..744
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5846098"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP E11"
 /notes="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Ecor I; Site_2: Not I; NCI CGAP E11 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTCAC.
 TAG TISSUE=chondrosarcoma
 TAG LIB=UI-H-E11
 TAG_SEQ=AACTGTCAC"

ORIGIN
 Query Match 11.1%; Score 52; DB 12; Length 744;
 Best Local Similarity 100.0%; Pred. No. 9.6e-15;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 373 CCTGTTCAATAGAAATGTTCCCTTCCCTGTTGTTCAATGTTGTAAGAT 424
 |||||
 Db 660 CCTGTTCAATAGAAATGTTCCCTTCCCTGTTGTTCAATGTTGTAAGAT 609
 |||||

RESULT 40
 BQ234959
 LOCUS
 DEFINITION BQ234959.g1 Canis Total Brain cDNAs Canis familiaris cDNA clone
 hd52b09 5', mRNA sequence.
 ACCESSION BQ234959
 VERSION BQ234959.1 GI:20430835
 KEYWORDS EST.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 574)
 O'Shaughnessy, A.L., Palmer, L., McCombie, W.R., Baker, J.P., Bahret, A., Cunniss, D., Dedhia, N., de la Bastide, M., Katzenberger, F., King, L., Kirchoff, K.A., Miller, B., Muller, S., Nascimento, L.U., Preston, R.R., Shah, R.S., Spiegel, L.A., Zutavern, T., Santos, L. and Hannon, G.J.
 Expressed sequence tags from Canis familiaris (dog) (5_2002)
 Unpublished (2002)
 Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884

Fax: 516 367 8874
Email: mccombie@cshl.org
Plate: hd52 row: b column: 09
Seq primer: -21M13UnivRev
High quality sequence stop: 574.
Location/Qualifiers
1. .574
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="hd52b09"
/clone_lib="Canis Total Brain cDNAs"
/note="Vector: Lambda Zap II; The library was provided by Greg Hannon and Lee Santos (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stratagene zap cDNA synthesis kit. It was made from dog whole brain cells. please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."

ORIGIN

Query Match 8.8%; Score 41; DB 13; Length 574;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TATGATCTTATCTACTAATTTATACAAAGATATTAAGGCCCT 375
|||||
Db 395 TATGATCTTATCTACTAATTTATACAAAGATATTAAGGCCCT 435
|||||

RESULT 41
LOCUS AI676745/c 491 bp mRNA linear EST 19-MAY-1999
DEFINITION etmEST0524 Eth1 Eimeria tenella cDNA clone etmJ085 5', mRNA
sequence.
ACCESSION AI676745.1 GI:4877225
VERSION AI676745
KEYWORDS EST.
SOURCE Eimeria tenella
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.

REFERENCE 1 (bases 1 to 491)
AUTHORS Wan,K.L., Chong,S.P., Ng,S.T., Tomley,F.M. and Jangji,M.S.
TITLE Survey of genes in Eimeria tenella merozoite
JOURNAL Unpublished (1999)
COMMENT Contact: Wan KL
Centre for Gene Analysis and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor DE, Malaysia
Tel: 6 03 8292997
Fax: 6 03 8293249
Email: klawan@pkriscc.cc.ukm.my
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: SK.
Location/Qualifiers
1. .491
/organism="Eimeria tenella"
/mol_type="mRNA"
/strain="Houghton"
/db_xref="taxon:5802"
/clone="etmJ085"
/dev_stage="second generation merozoite"
/lab_host="XLI-Blue MRF"
/clone_lib="Eth1"
/note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI;
Second generation merozoites of E. tenella H were purified by column chromatography and mRNA extracted using a FASTTRACK kit (Invitrogen). cDNA was synthesised and a Uni-ZAP XR library was constructed using cDNA synthesis kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III Gold cloning kit (Stratagene). The library was amplified

FEATURES
source

once through E. coli XLI-Blue MRF'."

ORIGIN

Query Match 7.3%; Score 34; DB 9; Length 491;
Best Local Similarity 100.0%; Pred. No. 9e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 TGCTGCTAACGCTGCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 486 TGCTGCTAACGCTGCTGCTGCTGCTGCTGCTGCT 453
|||||

RESULT 42
LOCUS CNS05DCI 972 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 042N02 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL332235
VERSION AL332235.1 GI:8225993
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE 1
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 972)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.
FEATURES
Location/Qualifiers
1. .972
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="042N02"
/clone_lib="A"
/note="Genoscope sequence ID : COAA042DG01A1-end : T3"

ORIGIN

Query Match 6.4%; Score 30; DB 29; Length 972;
Best Local Similarity 100.0%; Pred. No. 0.00098;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GCTAACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 687 GCTAACGCTGCTGCTGCTGCTGCTGCTGCT 716
|||||

```

RESULT 43
AA363384
LOCUS
DEFINITION
  ES732327 Ovary I Homo sapiens cDNA 5' end similar to H. sapiens hypothetical protein CTG-B33 (GB:LI0376), mRNA sequence.
ACCESSION
AA363384
VERSION
AA363384.1 GI:2015704
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Rhodes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Xu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL
Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE
96026280
PUBMED
7566098
COMMENT
Other ESTs: THC172576
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/html)
Seq primer: M13 Reverse
FEATURES
    source
        1..311
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="ATCC (inhost):167640"
        /db_xref="taxon:9606"
        /sex="female"
        /dev_stage="adult, 71 yrs"
        /clone_lib="Ovary I"
        /note="Organ: ovary; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN
Query Match 6.2%; Score 29; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 195 GCTGCTGCTGCTGCTGCTGCTGCTTAAAG 223
FEATURES
    source
        1..311
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="ATCC (inhost):167640"
        /db_xref="taxon:9606"
        /sex="female"
        /dev_stage="adult, 71 yrs"
        /clone_lib="Ovary I"
        /note="Organ: ovary; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN
Query Match 6.2%; Score 29; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 195 GCTGCTGCTGCTGCTGCTGCTGCTTAAAG 223

```

```

DEFINITION
f06500r Testis 5 Homo sapiens cDNA clone f06500 3' end, mRNA sequence.
ACCESSION
AA065331
VERSION
AA065331.1 GI:1929211
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 405)
Guellaen,G.
Guellaen,G. Unpublished (1996)
TITLE
Unpublished (1996)
JOURNAL
COMMENT
Contact: Guellaen G
Unite INSERM 99
INSERM
Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
Tel: (33)149813530
Fax: (33)14980908
Email: guellaen@infobiogen.fr
This sequence derives from a clone which was selected from the cDNA library - Testis 5 - using a repeat of 14 CAG as probe
Seq primer: M13 reverse
FEATURES
    Location/Qualifiers
        1..405
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="f06500"
        /clone_lib="Testis 5"
        /note="Vector: pSPORT1; Site 1: MluI; Site 2: NotI; mRNA was prepared from human testis of a 27 years old man. cDNA was prepared using a 15mer oligo dt anchored by two degenerated bases at its 3' end and containing a NotI site at its 5' end. The cDNA was cloned and containing a NotI site sites of pSPORT1. The MluI-SalI fragment come from the adaptor used for the cloning. The 3' end is at the NotI site. cDNA corresponding to abundant species were eliminated from this library."
ORIGIN
Query Match 6.2%; Score 29; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 281 GCTGCTGCTGCTGCTGCTGCTGCTTAAAG 309
RESULT 45
LBAP096B01
LOCUS
DEFINITION
Leishmania braziliensis GSS, clone LBAP096B01, genomic survey sequence.
ACCESSION
BX541495
VERSION
BX541495.1 GI:32139660
KEYWORDS
GSS; genomic survey sequence.
SOURCE
Leishmania braziliensis
ORGANISM
Leishmania braziliensis
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania; Leishmania braziliensis species complex.
REFERENCE
1
Laurentino,E.C., Ruiz,J.C. and Cruz,A.K.
GSS analysis of the Leishmania braziliensis genome
Unpublished
TITLE
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 423)
Cruz,A.K.
Direct Submission
TITLE
Direct Submission
JOURNAL
COMMENT
Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo, Department of Molecular and Cell Biology, FMRP, Avenida Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
Clone requests: akcruz@fmrp.usp.br.

```

```

FEATURES
  source
    1..423
      /organism="Leishmania braziliensis"
      /mol_type="genomic DNA"
      /strain="MHOM/BR/75/M2904"
      /db_xref="taxon:5660"
      /clone="LBAF096B01"

ORIGIN
Query Match          6.2%; Score 29; DB 29; Length 423;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 CTAACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 180 CTAACGCTGCTGCTGCTGCTGCTGCTGCT 208

RESULT 46
LOCUS T58423          482 bp      mRNA      linear      EST 09-FEB-1995
DEFINITION Yb60h09.r1 Stragatene ovary (#937217) Homo sapiens cDNA clone
IMAGE:75617 5', mRNA sequence.
ACCESSION T58423
VERSION T58423.1 GI:660260
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevasakis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 613
High quality sequence stops: 224
Clone is available royalty-free through LNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 224.

FEATURES
  source
    1..482
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="GDB:505346"
      /db_xref="taxon:9606"
      /clone="IMAGE:75617"
      /sex="female"
      /dev_stage="49 year old"
      /lab_host="SOLR cells (kanamycin resistant)"
      /clone_lib="Stratagene ovary (#937217)"
      /note="Organ: ovary; Vector: Bluescript SK; Site 1: EcoRI;
      Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
      Total ovary tissue, normal, caucasian. Average insert
      size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
      GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
      CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Query Match          6.2%; Score 29; DB 12; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 473 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 501

RESULT 48
LOCUS BG528108      855 bp      mRNA      linear      EST 03-APR-2001
DEFINITION BG528108F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:466572 5',
mRNA sequence.
ACCESSION BG528108
VERSION BG528108.1 GI:13519645
KEYWORDS EST.

FEATURES
  source
    1..732
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:4547111"
      /tissue_type="adenocarcinoma cell line"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_15"
      /note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dT priming. Directionally
      cloned into EcoRI/XhoI sites using the following 5',
      adaptor: GGCACGAG(G). Size-selected >500bp for average
      insert size 1.8kb. Library constructed by Ling Hong in
      the laboratory of Gerald M. Rubin (University of
      California, Berkeley) using ZAP-cDNA synthesis kit
      (Stratagene) and Superscript II RT (Life Technologies)"

ORIGIN
Query Match          6.2%; Score 29; DB 12; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 473 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 501

RESULT 48
LOCUS BG528108      855 bp      mRNA      linear      EST 03-APR-2001
DEFINITION BG528108F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:466572 5',
mRNA sequence.
ACCESSION BG528108
VERSION BG528108.1 GI:13519645
KEYWORDS EST.

```

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 855)
JOURNAL     NIH-MGC http://mgc.nci.nih.gov/.
COMMENT     National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILNL at:
            http://image.llnl.gov
            Plate: LUCM1497 row: 0 column: 21
            High quality sequence stop: 594.
FEATURES    Location/Qualifiers
            1..855
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4686572"
                /tissue_type="mucoepidermoid carcinoma"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_lib="NIH_MGC_59"
                /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
                SfiI (ggccgccctggcc); Site 2: SfiI (ggccattatggcc);
                Double-stranded cDNA was prepared from cell line RNA. 5'
                and 3' adaptors were used in cloning as follows: 5'
                adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
                sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dT(30)BN-3'
                (where B = A, C, or G and N = A, C, G, or T). Average
                insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                contained inserts by PCR. This library was enriched for
                full-length clones and was constructed by Clontech
                Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                Library."
ORIGIN
Query Match      6.2%; Score 29; DB 12; Length 855;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  54  GCTGCTGCTGCTGCTGCTGCTTAAG 82
Db    483  GCTGCTGCTGCTGCTGCTGCTTAAG 511

RESULT 49
LOCUS    BJ600203/c
DEFINITION BJ600203 323 bp mRNA linear EST 22-OCT-2003
            Bu600203 normalized full length cDNA library, chloronemata,
            caulonemata and rhizoid-like protonemata Physcomitrella patens
            subsp. patens cDNA clone pphn28014 3', mRNA sequence.
ACCESSION BJ600203
VERSION    BJ600203.1 GI:37842195
KEYWORDS   EST.
SOURCE     Physcomitrella patens subsp. patens
ORGANISM   Physcomitrella patens subsp. patens
            Bukaryota; Viridiplantae; Streptophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE  1 (bases 1 to 323)
AUTHORS     Nishiyama,T., Fujita,T., Shin-i,T., Seki,M., Nishide,H.,
            Uchiyama,I., Kamiya,A., Carninci,P., Hayashizaki,Y., Shinozaki,K.,
            Kohara,Y. and Hasebe,M.
TITLE       Comparative genomics of Physcomitrella patens gametophytic
            transcription and Arabidopsis thaliana: implication for land plant
            evolution
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
MEDLINE    22709184

12808149
COMMENT    Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            A backbone of the vector is basically from pBluescript II (KS),
            that was in vivo excised from a 1-FLC phage vector (Carninci et al.
            2001). 5' end of the cDNA that was digested with XhoI was ligated
            to SalI site of the vector and the 3' end including polyA tail was
            ligated to BamHI site of the
            vector(5'- gagAgAgAgAgATCAACCCCTggAgAgTTTTTTTTTTTTTN-3' was
            used as a 1st 3' primer, and
            5'-ggTTCgAgTCATcGTTCAGACgATgACTCGAGACCGNNNN-3' as 2nd
            5'-hairpin primer, giving the following 5' boarder sequence,
            AGCCAAATCGCCGAGCTCGAATTCGTCGAGAACCG). cDNA instert could be
            amplified with conventional T7 and T3 primers. This full-length
            cDNA library was generated according to the method described in
            Nishiyama et al. (2003).
            Protonemata were blerded by the POLYTRON, and then cultivated on
            the BCDATG medium for 13- 14 days under the continuous light.
            These clones are available from RIKEN Bio Resource Center
            (http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database
            of Physcomitrella EST clones is available at the PHYSCObase
            (http://moss.nibb.ac.jp).
FEATURES    Location/Qualifiers
            1..323
                /organism="Physcomitrella patens subsp. patens"
                /mol_type="mRNA"
                /sub_species="patens"
                /db_xref="taxon:145481"
                /clone="pphn28014"
                /tissue_type="mixture of chloronemata, caulonemata and
                rhizoid-like protonemata"
                /clone_lib="normalized full length cDNA library,
                chloronemata, caulonemata and rhizoid-like protonemata"
ORIGIN
Query Match      6.0%; Score 28; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  52  ACGCTGCTGCTGCTGCTGCTGCTGCTTA 79
Db    278  ACGCTGCTGCTGCTGCTGCTGCTTA 251

RESULT 50
LOCUS    AW509816
DEFINITION AW509816 345 bp mRNA linear EST 03-MAR-2000
            gas6a05.y1 Moss EST library PPU Physcomitrella patens cDNA clone
            PEP SOURCE ID:PPU110509 5', similar to TR:Q9ZVB7 Q9ZVB7 PUTATIVE
            STRICTOSIDINE SYNTHASE. ;, mRNA sequence.
ACCESSION AW509816
VERSION    AW509816.1 GI:7147894
KEYWORDS   EST.
SOURCE     Physcomitrella patens
ORGANISM   Physcomitrella patens
            Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE  1 (bases 1 to 345)
AUTHORS     Quatrano,S., Barra,M., Hallier,B., Pape,D., Martin,J., Wylie,T.,
            Underwood,K., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter,E.,
            Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE       Leeds/Wash U Moss EST Project
JOURNAL     Unpublished (1999)
COMMENT    Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine

```

Wed May 26 09:35:15 2004

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Libraries were constructed by Dr. Stavros Bashirides as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco

High quality sequence stop: 293.

FEATURES

source

```

1. 345
Location/Qualifiers
/organism="Physcomitrella patens"
/mol_type="mRNA"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID:PPU110509"
/tissue_type="protonemata: 7 day old tissue ammonium-grown"
/lab_host="DH10B"
/clone_lib="Moss EST library PPU"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Construction of the cDNA library was carried out using Stratagene's UniZAP - cDNA synthesis kit'. cDNA was constructed using an oligo dt primer/linker that contains a XhoI site within it. Following ds cDNA synthesis, EcoRI adapters were ligated to the blunt ends and sample was digested with XhoI. The result is cDNA with an EcoRI sticky end on one side and a XhoI sticky end on the other. This cDNA was ligated directionally in UniZAP arms. The vector is designed containing the pBluescript sequence as well as lambda DNA and cDNA is cloned within this pBluescript sequence. The vector was then packaged using Gold giga packaging extracts. Library was grown in XL1Blue MRF' cells and amplified. The library was excised by mass excision using Stratagene's Mass excision kit' that uses exassit as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids. SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered by using Quiagen Midi prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation."

```

ORIGIN

Query Match 6.0%; Score 28; DB 10; Length 345;
 Best Local Similarity 100.0%; Pred. No. 0.0086;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 52 ACCTGCTGCTGCTGCTGCTGCTTA 79
   |||||
Db 295 ACCTGCTGCTGCTGCTGCTGCTTA 322

```

Search completed: May 25, 2004, 19:19:03
 Job time : 2889 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 15:14:46 ; Search time 95 Seconds
(without alignments)
2728.022 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctatccacct.....taataaacagttaaagctg 467

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| C 1 | 467 | 100.0 | 50000 | 4 | US-09-146-053-4 |
| C 2 | 40.4 | 8.7 | 7218 | 1 | US-08-232-463-14 |
| C 3 | 38.2 | 8.2 | 4262 | 4 | US-09-521-511C-10 |
| C 4 | 37 | 7.9 | 2521 | 3 | US-09-115-446-1 |
| C 5 | 37 | 7.9 | 2521 | 1 | US-09-565-590-1 |
| C 6 | 37 | 7.9 | 2526 | 3 | US-09-115-446-5 |
| C 7 | 37 | 7.9 | 2526 | 4 | US-09-565-590-5 |
| C 8 | 36.4 | 7.8 | 3001 | 4 | US-09-539-333D-178 |
| C 9 | 35.8 | 7.7 | 800 | 1 | US-08-468-853-7 |
| C 10 | 35.8 | 7.7 | 800 | 1 | US-08-468-855-7 |
| C 11 | 35.8 | 7.7 | 800 | 1 | US-08-310-357-7 |
| C 12 | 35.8 | 7.7 | 800 | 1 | US-08-468-852-7 |
| C 13 | 35.8 | 7.7 | 800 | 2 | US-08-468-857-7 |
| C 14 | 35.6 | 7.6 | 107 | 4 | US-09-205-995-75 |
| C 15 | 35.6 | 7.6 | 169 | 4 | US-09-205-995-72 |
| C 16 | 35.6 | 7.6 | 190 | 4 | US-09-205-995-77 |
| C 17 | 35.4 | 7.6 | 430 | 4 | US-09-621-976-16656 |
| C 18 | 35.4 | 7.6 | 1853 | 1 | US-08-553-110-2 |
| C 19 | 35.4 | 7.6 | 4168 | 4 | US-09-266-225D-17 |
| C 20 | 35.4 | 7.6 | 4279 | 3 | US-09-041-886-22 |
| C 21 | 35.2 | 7.5 | 2160 | 3 | US-09-588-256-1 |
| C 22 | 35.2 | 7.5 | 3311 | 4 | US-09-367-891A-5 |
| C 23 | 35.2 | 7.5 | 9053 | 4 | US-09-976-594-306 |
| C 24 | 35.2 | 7.5 | 10807 | 1 | US-08-206-176-7 |
| C 25 | 35.2 | 7.5 | 10807 | 2 | US-08-756-506-5 |
| C 26 | 35.2 | 7.5 | 176373 | 3 | US-09-128-155-17 |
| C 27 | 35 | 7.5 | 10348 | 2 | US-08-457-273B-41 |

| | | | | | |
|-----|-----|---------|---|-------------------|-------------------|
| 35 | 7.5 | 10348 | 3 | US-08-556-419-13 | Sequence 13, Appl |
| 35 | 7.5 | 10348 | 3 | US-09-041-886-14 | Sequence 1, Appl |
| 35 | 7.5 | 10366 | 1 | US-08-246-982A-5 | Sequence 5, Appl |
| 35 | 7.5 | 10366 | 1 | US-08-453-265-5 | Sequence 5, Appl |
| 32 | 7.5 | 3819 | 1 | US-07-686-322A-3 | Sequence 3, Appl |
| 32 | 7.5 | 3819 | 1 | US-08-002-999-3 | Sequence 3, Appl |
| 33 | 7.4 | 325 | 2 | US-08-531-927B-3 | Sequence 3, Appl |
| 34 | 7.4 | 397 | 3 | US-09-253-691-3 | Sequence 3, Appl |
| 35 | 7.4 | 477 | 3 | US-09-135-994-1 | Sequence 1, Appl |
| 36 | 7.4 | 477 | 3 | US-09-684-843A-1 | Sequence 1, Appl |
| 37 | 7.4 | 477 | 3 | US-08-481-814A-3 | Sequence 3, Appl |
| 38 | 7.4 | 1332 | 2 | US-08-836-582-1 | Sequence 871, App |
| 39 | 7.4 | 1489 | 3 | US-08-836-582-1 | Sequence 1, Appl |
| 40 | 7.4 | 1489 | 4 | US-09-242-737-3 | Sequence 1, Appl |
| 41 | 7.4 | 1489 | 4 | US-09-023-655-871 | Sequence 3, Appl |
| 42 | 7.4 | 1489 | 4 | US-08-531-927B-1 | Sequence 1, Appl |
| 43 | 7.4 | 1776 | 2 | US-09-041-886-12 | Sequence 12, Appl |
| 44 | 7.4 | 1776 | 3 | US-08-479-913E-1 | Sequence 1, Appl |
| 45 | 7.4 | 6177 | 4 | US-09-091-042A-1 | Sequence 1, Appl |
| 46 | 7.4 | 7257 | 4 | US-09-146-053-4 | Sequence 4, Appl |
| 47 | 7.4 | 50000 | 4 | US-08-916-421B-1 | Sequence 1, Appl |
| 48 | 7.4 | 1664976 | 4 | US-08-545-528D-1 | Sequence 1, Appl |
| 49 | 7.3 | 580073 | 4 | US-09-135-994-3 | Sequence 3, Appl |
| 50 | 7.3 | 300 | 3 | US-09-684-843A-3 | Sequence 3, Appl |
| 51 | 7.3 | 300 | 4 | US-08-184-009-150 | Sequence 150, App |
| 52 | 7.3 | 510 | 2 | US-08-458-356-150 | Sequence 150, App |
| 53 | 7.3 | 510 | 2 | US-08-460-736-150 | Sequence 150, App |
| 54 | 7.3 | 510 | 3 | US-09-535-370-150 | Sequence 2, Appl |
| 55 | 7.3 | 510 | 4 | US-09-634-368-2 | Sequence 2, Appl |
| 56 | 7.3 | 1472 | 4 | US-09-634-368-11 | Sequence 11, Appl |
| 57 | 7.3 | 2013 | 4 | US-08-480-547A-11 | Sequence 11, Appl |
| 58 | 7.3 | 2060 | 1 | US-08-250-847B-11 | Sequence 11, Appl |
| 59 | 7.3 | 2060 | 1 | US-08-463-949A-11 | Sequence 11, Appl |
| 60 | 7.3 | 2060 | 2 | US-08-464-410A-11 | Sequence 11, Appl |
| 61 | 7.3 | 2060 | 3 | PCT-US94-06066-11 | Sequence 11, Appl |
| 62 | 7.3 | 2060 | 5 | US-08-480-547A-22 | Sequence 22, Appl |
| 63 | 7.3 | 2645 | 1 | US-08-480-547B-22 | Sequence 22, Appl |
| 64 | 7.3 | 2645 | 2 | US-08-463-949A-22 | Sequence 22, Appl |
| 65 | 7.3 | 2645 | 3 | US-08-464-410A-22 | Sequence 22, Appl |
| 66 | 7.3 | 2645 | 5 | PCT-US94-06066-22 | Sequence 22, Appl |
| 67 | 7.3 | 2709 | 4 | US-09-634-368-13 | Sequence 13, Appl |
| 68 | 7.3 | 2709 | 4 | US-08-948-113D-24 | Sequence 24, Appl |
| 69 | 7.2 | 3215 | 4 | US-08-320-559-29 | Sequence 29, Appl |
| 70 | 7.2 | 3376 | 3 | US-08-545-860D-29 | Sequence 29, Appl |
| 71 | 7.2 | 3376 | 3 | PCT-US94-04496-29 | Sequence 29, Appl |
| 72 | 7.2 | 3376 | 5 | US-08-469-802B-4 | Sequence 4, Appl |
| 73 | 7.2 | 168 | 2 | US-08-267-803B-4 | Sequence 4, Appl |
| 74 | 7.2 | 168 | 2 | US-08-469-802B-5 | Sequence 5, Appl |
| 75 | 7.2 | 171 | 1 | US-08-267-803B-5 | Sequence 5, Appl |
| 76 | 7.2 | 171 | 1 | US-08-469-802B-2 | Sequence 2, Appl |
| 77 | 7.2 | 195 | 1 | US-08-469-802B-2 | Sequence 2, Appl |
| 78 | 7.2 | 195 | 2 | US-08-267-803B-2 | Sequence 3, Appl |
| 79 | 7.2 | 234 | 1 | US-08-469-802B-3 | Sequence 3, Appl |
| 80 | 7.2 | 234 | 2 | US-08-267-803B-3 | Sequence 3, Appl |
| 81 | 7.2 | 1037 | 4 | US-09-181-585-3 | Sequence 3, Appl |
| 82 | 7.2 | 1159 | 4 | US-09-181-585-1 | Sequence 2, Appl |
| 83 | 7.2 | 1471 | 4 | US-09-181-585-2 | Sequence 2, Appl |
| 84 | 7.2 | 2892 | 1 | US-08-264-534-5 | Sequence 5, Appl |
| 85 | 7.2 | 2892 | 1 | US-08-083-590A-1 | Sequence 1, Appl |
| 86 | 7.2 | 2892 | 1 | US-08-465-500-5 | Sequence 5, Appl |
| 87 | 7.2 | 2892 | 2 | US-08-346-126-5 | Sequence 5, Appl |
| 88 | 7.2 | 2892 | 2 | US-08-346-128-5 | Sequence 5, Appl |
| 89 | 7.2 | 2892 | 3 | US-08-532-384-1 | Sequence 1, Appl |
| 90 | 7.2 | 2892 | 3 | US-08-893-828-5 | Sequence 7, Appl |
| 91 | 7.2 | 2950 | 5 | PCT-US93-08386-7 | Sequence 7, Appl |
| 92 | 7.2 | 2968 | 5 | PCT-US93-08386-1 | Sequence 1, Appl |
| 93 | 7.2 | 3366 | 2 | US-08-469-802B-1 | Sequence 1, Appl |
| 94 | 7.2 | 3366 | 2 | US-08-267-803B-1 | Sequence 1, Appl |
| 95 | 7.2 | 7218 | 1 | US-08-232-463-14 | Sequence 14, Appl |
| 96 | 7.2 | 10660 | 3 | US-08-267-803B-8 | Sequence 8, Appl |
| 97 | 7.2 | 10660 | 3 | US-09-041-886-16 | Sequence 16, Appl |
| 98 | 7.2 | 270 | 4 | US-09-146-054-8 | Sequence 8, Appl |
| 99 | 7.2 | 270 | 4 | US-09-664-977A-8 | Sequence 8, Appl |
| 100 | 7.2 | 506 | 1 | US-08-469-802B-7 | Sequence 7, Appl |

```

; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match      8.7%; Score 40.4; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 0.034;
Matches 5; Conservative 134; Mismatches 75; Indels 0; Gaps 0;

QY      6 TTGCGCTATCCACCTCCCAAGCGCTCATGCTGGAGTGGGACTGGTGGTCCGACGAGAAAGTCTC 125
Db      1229 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1288
QY      66 GCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTCCGACGAGAAAGTCTC 125
Db      1289 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1348
QY      126 TTGCGCACTGAGCGCCCATCAGGATGGGCTTCTTCCGCCCTTCTTCCCTTCTTCTGCT 185
Db      1349 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1408
QY      186 CTGCGCTCATGCGCTGCGCATGACCTGACGACCA 219
Db      1409 YYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAA 1442

RESULT 3
US-09-521-511C-10/c
; Sequence 10, Application US/09521511C
; Patent No. 655358
; GENERAL INFORMATION:
; APPLICANT: Gurnett, Anne
; APPLICANT: Liberator, Paul A.
; APPLICANT: Donald, Robert
; APPLICANT: Schmatz, Dennis
; APPLICANT: Harris, Georgianna
; APPLICANT: Rattray, Sandra J.
; TITLE OF INVENTION: CYCLIC GMP DEPENDENT PROTEIN KINASE AS A

```

```

; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match      100.0%; Score 467; DB 4; Length 50000;
Best Local Similarity 100.0%; Pred. No. 1.15e-136;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAATGTTTGCCTATCCACCTCCCAAGCGCTTACCTATGCTGCTGCTAAAGCTGCTG 60
Db      1518 CAATGTTTGCCTATCCACCTCCCAAGCGCTTACCTATGCTGCTGCTAAAGCTGCTG 1459
QY      61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGCTGCTGCTGCTG 120
Db      1458 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGCTGCTGCTGCTG 1399
QY      121 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Db      1398 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1339
QY      181 TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db      1338 TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279
QY      241 GAGAAAGTGGGGATGCTAAGAAAGCTGAGAGTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 300
Db      1278 GAGAAAGTGGGGATGCTAAGAAAGCTGAGAGTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 1219
QY      301 GGCTAGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db      1218 GGCTAGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1159
QY      361 AGATATTAAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db      1158 AGATATTAAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1099
QY      421 AGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467
Db      1098 AGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1052

RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS

```

| | Query Match | 7.9% | Score 37; | DB 3; | Length 2521; |
|----|-----------------------|------------------|---|------------|--------------|
| | Best Local Similarity | 64.7%; | Pred. No. 0.21; | | |
| | Matches | 55; Conservative | 0; Mismatches | 30; Indels | 0; Gaps |
| QY | 41 | TGTCGTCTAA | CGCTCTCTGCTGCTGCTGCTTAAAGGCTCATGCTGAGTGGG | 100 | |
| DB | 408 | TGTCGTCTGTT | GCTCTGCTGCTGCTGCTGCTCATCCACAGGATGCACGGG | 349 | |
| QY | 101 | GACTTGTGCTG | CGCCAGAAAGTCTC | 125 | |
| DB | 348 | CACTTGGGTCTT | CATCAAGTCCC | 324 | |

RESULT 6
US-09-115-446-5/c
Sequence 5, Application US/09115446
Patent No. 6165719
GENERAL INFORMATION:
APPLICANT: Chandy, George K.
APPLICANT: Gargus, Jay J.
APPLICANT: Gutman, George
APPLICANT: Fantino, Emmanuelle
APPLICANT: Kalman, Katarin
TITLE OF INVENTION: hKCA3/KCN3 SMALL CONDUCTANCE CALCIUM
TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
FILE REFERENCE: 07306/014001
CURRENT APPLICATION NUMBER: US/09/115.446
CURRENT FILING DATE: 1998-07-14
EARLIER APPLICATION NUMBER: 60/052,556
EARLIER FILING DATE: 1997-07-15
EARLIER APPLICATION NUMBER: 60/070,741
EARLIER FILING DATE: 1998-01-08
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 2526
TYPE: DNA
ORGANISM: Homo sapiens
US-09-115-446-5

Query Match


```

; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5670362el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,853
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-Jun-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; CLONE: Bam20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-853-7

```

```

Query Match 7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 6 TTTCGCTATCCACTCCCAAGCCCTTACCTATGCTGCTGCTAAAGCTGCTGCT 65
Db 461 TTTTCGCTCAACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402

Qy 66 GCTGCTGCTGCTTAAAGCTCATGCTT 92
Db 401 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375

```

```

RESULT 10
US-08-468-855-7/c
; Sequence 7, Application US/08468855
; Patent No. 5780289
; GENERAL INFORMATION:
; APPLICANT: van den BOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE

```

```

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5780289el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,855
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-Jun-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; CLONE: Bam20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-855-7

```

```

Query Match 7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 6 TTTCGCTATCCACTCCCAAGCCCTTACCTATGCTGCTGCTAAAGCTGCTGCT 65
Db 461 TTTTCGCTCAACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402

Qy 66 GCTGCTGCTGCTTAAAGCTCATGCTT 92
Db 401 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375

```

```

RESULT 11
US-08-310-357-7/c
; Sequence 7, Application US/08310357
; Patent No. 5789233
; GENERAL INFORMATION:
; APPLICANT: van den BOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Organon Teknika Corporation
; ADDRESSEE: Biotechnology Research Institute
; STREET: 1330-A Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310,357
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/904,075
; FILING DATE:
; FILING DATE:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-Jun-1991
; NAME: William M. Blackstone
; REGISTRATION NUMBER: 29,772
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; IMMEDIATE SOURCE:
; CLONE: Em20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-310-357-7

```

```

Query Match          7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      6 TTGGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCT 65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      461 TTTCGCTCCCACTGCAGCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65

QY      66 GCTGCTGCTGCTTAAAGGCTCATGCTT 92
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      401 GCTGCTGCTGCTGCCATGGTTCTTCAT 375

```

```

RESULT 12
US-08-468-852-7/c
; Sequence 7, Application US/08468852
; Patent No. 5792644
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5792644el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,852
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-Jun-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gornley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; IMMEDIATE SOURCE:
; CLONE: Em20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-852-7

```

```

Query Match          7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      6 TTGGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCT 65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      461 TTTCGCTCCCACTGCAGCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65

QY      66 GCTGCTGCTGCTTAAAGGCTCATGCTT 92
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      401 GCTGCTGCTGCTGCCATGGTTCTTCAT 375

```

```

RESULT 13
US-08-468-857-7/c
; Sequence 7, Application US/08468857
; Patent No. 5925347
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5925347el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,857
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; CLONE: Ham20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-857-7
```

```
Query Match 7.7%; Score 35.8; DB 2; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 32;

QY 6 TTTCCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTAAAGCTGCTGCTGCT 65
Db 461 TTTCCTCCCACTGAGTGTGCAGTGCAGTGCAGTGCAGTGCAGTGTGCTGCT 402

QY 66 GCTGCTGCTGCTTAAAGGCTCATGCTT 92
Db 401 GCTGCTGCTGCTGCCATGGTCTTCAT 375
```

```
RESULT 14
US-09-205-995-75/c
; Sequence 75, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
; PRIOR APPLICATION NUMBER: 08/661,627
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 107
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Description of Artificial Sequence: Reverse gene
; OTHER INFORMATION: construct corresponding to a specific region of
; OTHER INFORMATION: the mouse Ii gene.
US-09-205-995-75
```

```
Query Match 7.6%; Score 35.6; DB 4; Length 107;
Best Local Similarity 71.2%; Pred. No. 0.079; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 19;

QY 20 TCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTGCTTAA 79
Db 73 TCCACAGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14

QY 80 AAGGCT 85
Db 13 GTGTCT 8
```

```
RESULT 15
US-09-205-995-72/c
; Sequence 72, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
; PRIOR APPLICATION NUMBER: 08/661,627
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse gene
; OTHER INFORMATION: construct corresponding to a specific region of
; OTHER INFORMATION: the mouse Ii gene.
US-09-205-995-72
```

```
Query Match 7.6%; Score 35.6; DB 4; Length 169;
Best Local Similarity 71.2%; Pred. No. 0.11; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 19;

QY 20 TCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTGCTTAA 79
Db 73 TCCACAGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14

QY 80 AAGGCT 85
Db 13 GTGTCT 8
```

```
RESULT 16
US-09-205-995-77/c
; Sequence 77, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
```



```

/ APPLICANT: Burke, James R.
/ APPLICANT: Vance, Jeffrey M.
/ APPLICANT: Engchild, Jan
/ APPLICANT: Strittmatter, Warren J.
/ TITLE OF INVENTION: Therapeutics for Diseases Associated
/ TITLE OF INVENTION: with CAG Trinucleotide Repeat Expansion
/ NUMBER OF SEQUENCES: 2
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Kenneth D. Sibley
/ STREET: P.O. Drawer 34009
/ CITY: Charlotte
/ STATE: No. 5723301th Carolina
/ COUNTRY: United States of America
/ ZIP: 28234
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/553,110
/ FILING DATE:
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Sibley, Kenneth D.
/ REGISTRATION NUMBER: 31,665
/ REFERENCE/DOCKET NUMBER: 5405-117
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 919-420-2200
/ TELEFAX: 919-881-3175
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1853 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/
US-08-553-110-2

Query Match          7.6%; Score 35.4; DB 1; Length 1853;
Best Local Similarity 66.2%; Pred No. 0.55; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 26;

QY      41 TCCTGCTGTCAACGTCGTGCTGCTGCTGCTTAAGAGCTCATGCTTGAGTGGG 100
Db      1244 TCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGATGCTGTGAGACTGCT 1185

QY      101 GACTGGTCGGTGCCCGAG 117
Db      1184 GGGTGGCGGTGGACTG 1168

RESULT 19
/ Sequence 17, Application US/09266225D
/ Patent No. 6573364
/ GENERAL INFORMATION:
/ APPLICANT: Nandabalan, Krishan
/ APPLICANT: Kingsmore, Stephen
/ APPLICANT: Tchervet, Velizar
/ TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak
/ TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Protein-
/ TITLE OF INVENTION: Interacting Proteins
/ FILE REFERENCE: 15966-523
/ CURRENT APPLICATION NUMBER: US/09/266,225D
/ CURRENT FILING DATE: 1999-03-10
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 17
/ LENGTH: 4168
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/
US-09-266-225D-17

```



```
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 306
; LENGTH: 9053
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 898877.6
; NAME/KEY: unsure
; LOCATION: 2006, 2012
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-306

Query Match          7.5%; Score 35.2; DB 4; Length 9053;
Best Local Similarity 54.9%; Pred. No. 1.7;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      8 TGCTATTCACCTCCGCCAAGCCCTTTTACCTATGCTGCTGCTAAAGCTGCTGCTGC 67
Db      2090 TTTCTATCTTTCTCCCTCCCTGCTGCGGGGATGGGCTGCTGCTGCTGCTGC 2031

QY      68 TGCTGCTGCTTAAAGCTCATGCTTGAGTGGGGAGTGTGGTCCCGAAGAGTCTTT 127
Db      2030 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1971

QY      128 CT 129
Db      1970 CT 1969

RESULT 24
US-08-206-176-7
; Sequence 7, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dalrymple, Michael A
; APPLICANT: Prunkard, Donna E
; APPLICANT: Foster, Donald C
; TITLE OF INVENTION: Production of Fibrinogen in Transgenic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,176
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-15
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-547-8080 ext 322
; TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ovine beta-lactoglobulin
US-08-206-176-7

Query Match          7.5%; Score 35.2; DB 1; Length 10807;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY      129 TGCCACTGAGCGCCCATCAGGGATTGGCCCTTTCTTCCCTTCTCTTCTGTCCT 188
Db      8265 TGACTCTCCCTCCCTCCACAGGCACTGCCACTTAGGTGAGCCCTGCCGGTGCTCTG 8324

QY      189 GCCTCATCGGCTGCCATGACCTGCAGCCCAAGCCCGCCGAGGGAGGAGAAAGT 248
Db      8325 GGGTAAGCTGCTGCTGCCCTGCCACGTCCTGGGCACACACATGGGGGTCTTGCT 8384

QY      249 GGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTAGGG 308
Db      8385 GGGGCTGGGAGCCCAACATCATGAGCCCTGGGGTCCCTGTGAGAAATGGCTGGAAGCTGG 8444

QY      309 GGGC 312
Db      8445 GGTG 8448

RESULT 25
US-08-756-506-5
; Sequence 5, Application US/08756506
; Patent No. 5905185
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,506
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 95-28
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```

1 APPLICANT: Lin, Biaoyang
2 APPLICANT: Nasir, Jamal
3 TITLE OF INVENTION: Mouse Model for Huntington's Disease and
4 TITLE OF INVENTION: Related DNA Sequences
5 NUMBER OF SEQUENCES: 42
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Virginia Bennett
8 STREET: PO Box 37428
9 CITY: Raleigh
10 STATE: No. 584995th Carolina
11 COUNTRY: US
12 ZIP: 27627
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/457,273B
20 FILING DATE:
21 CLASSIFICATION: 800
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Bennett, Virginia C.
24 REGISTRATION NUMBER: 37,092
25 REFERENCE/DOCKET NUMBER: 3477-85A
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 919-854-1400
28 TELEFAX: 919-854-1401
29 INFORMATION FOR SEQ ID NO: 41:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 10348 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: cDNA
36 US-08-457-273B-41
37
38 Query Match 7.5%; Score 35; DB 2; Length 10348;
39 Best Local Similarity 88.4%; Pred. No. 2.2;
40 Matches 38; Conservative 0; Mismatches 5; Indels 0
41
42 QY 41 TCGTCTGCTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAGG 83
43
44 DB 404 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGGAAG 362
45
46
47 RESULT 28
48 US-08-556-419-13/c
49 Sequence 13, Application US/08556419C
50 Patent No. 603549
51 GENERAL INFORMATION:
52 APPLICANT: Ross, Christopher
53 APPLICANT: Li, Xiao-Jiang
54 APPLICANT: Li, Shi-Hua
55 APPLICANT: Sharp, Alan
56 APPLICANT: Lananan, Anthony
57 APPLICANT: Worley, Paul
58 APPLICANT: Snyder, Solomon
59 TITLE OF INVENTION: Huntingtin-associated protein
60 FILE REFERENCE: 01107.52271
61 CURRENT APPLICATION NUMBER: US/08/556,419C
62 CURRENT FILING DATE: 1995-11-09
63 NUMBER OF SEQ ID NOS: 25
64 SOFTWARE: FastSeq for Windows Version 3.0
65 SEQ ID NO 13
66 LENGTH: 10348
67 TYPE: DNA
68 ORGANISM: Homo sapiens
69 US-08-556-419-13
70
71 Query Match 7.5%; Score 35; DB 3; Length 10348;
72 Best Local Similarity 88.4%; Pred. No. 2.2;
73 Matches 38; Conservative 0; Mismatches 5; Indels 0

```



```

; SEQ ID NO 1
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-684-843A-1

Query Match          7.4%; Score 34.4; DB 4; Length 477;
Best Local Similarity 60.9%; Pred.No. 0.49; 36; Indels 0; Gaps 0;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY      41   TGTGCTGTAAAGCTCTGCTGCTGTGCTGTGCTTTAAAGGCTCATGCTTGAGTGGG 100
Db       161   TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCGGGCGGCGGCTGCTCGCCGCC 102

QY      101   GACTGTCGTGCCCCACAAAAGTCTCTTCTGCC 132
Db       101   GCCGCGCCGCGCGGCGCGGCTCCCCTGCAC 70

RESULT 38
US-08-481-814A-3/C
; Sequence 3, Application US/08481814A
; Patent No. 5869040
; GENERAL INFORMATION:
; APPLICANT: Hsu, Yen-Ming
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Biogen, Inc.
; STREET: 14 Cambridge Center
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02142
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,814A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan, Warren A
; REFERENCE/DOCKET NUMBER: A001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-679-2000
; TELEFAX: 617-679-2838
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 63..1301
; OTHER INFORMATION: /function= "transcription factor"
; OTHER INFORMATION: /product= "E2F-4"
US-08-481-814A-3

Query Match          7.4%; Score 34.4; DB 2; Length 1332;
Best Local Similarity 86.4%; Pred.No. 0.93; 6; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      34   TTACCTATGCTGCTGAACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77

```

```

; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,566
; FILING DATE: 10-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,582
; FILING DATE: <Unknown>
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-265-566-1

Query Match 7.4%; Score 34.4; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 1;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 34 TTACTATGCTGCTGAACGCTGCTGCTGCTGCTGCTGCTGCT 77
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 991 TTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948

RESULT 41
US-09-242-737-3/c
; Sequence 3, Application US/09242737
; Patent No. 6368809
; GENERAL INFORMATION:
; APPLICANT: BERNARDS, REN(
; TITLE OF INVENTION: E2F UBIQUITINATION DOMAIN, AND ASSAYS FOR
; INHIBITORS OF E2F UBIQUITINATION
;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,737
; FILING DATE: 23-Feb-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02293
; FILING DATE: 22-AUG-1997
; APPLICATION NUMBER: GB 9617697.9
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.

```

```

; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-63
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-737-3

Query Match 7.4%; Score 34.4; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 1;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 34 TTACTATGCTGCTGAACGCTGCTGCTGCTGCTGCTGCTGCT 77
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 991 TTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948

RESULT 42
US-09-023-655-871/c
; Sequence 871, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 871:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1061145

```

US-09-023-655-871

Query Match 7.4%; Score 34.4; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 1;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy

34 TTACCTATGCTGTCTAAGCTGCTGCTGCTGCTGCTGCTGCT 77
|||||

Dd

991 TTA CTGTT GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
|||||

RESULT 43

```

US-08-531-927B-1/c
; Sequence 1, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491

```

```
Query Match      7.4%; Score 34.4; DB 2; Length 1776;
Best Local Similarity 78.8%; Pred No. 1.1;
Matches 41: Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

| | | | | |
|----|--|-----|--|-----|
| Qy | | 26 | AAGCCCCCTTACCTATGCTGCTGTCTAACGCTGCTGCTGCTGCTGCTGCTGCT | 77 |
| Dβ | | 994 | AGGTCCCGGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 943 |

RESULT 44

RESULT 44
 US-09-041-886-12/c
 ; Sequence 12, Application US/09041886
 ; Patent No. 6235872
 ; GENERAL INFORMATION:
 ; APPLICANT: Bredesen, Dale E.

```

1  APPLICANT: Rabizadeh, Sharroz
2
3  TITLE OF INVENTION: Prooprototic Peptides, Dependence
4  TITLE OF INVENTION: Polypeptides and Methods of Use
5  NUMBER OF SEQUENCES: 72
6
7  CORRESPONDENCE ADDRESS:
8
9  ADDRESSEE: Campbell & Flores LLP
10 STREET: 4370 La Jolla Village Drive, Suite 700
11 CITY: San Diego
12 STATE: California
13 COUNTRY: United States
14 ZIP: 92122
15
16 COMPUTER READABLE FORM:
17
18 MEDIUM TYPE: Floppy disk
19
20 COMPUTER: IBM PC compatible
21
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: PatentIn Release #1.0, Version #1.25
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/09/041,886
26 FILING DATE:
27

```

```
Query Match          7.4%; Score 34.4; DB 3; Length 1776;
Best Local Similarity 78.8%; Pred. No. 1.1;
Matches 41; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

QY 26 AAGCCCCCTTTACCTATGCTGCTGCTAAGCTGCTGCTGCTGCTGCTGCT 77

pb 994 AGGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943

RESULT 45

```

US-08-479-913E-1/c
; Sequence 1, Application US/08479913E
; Patent No. 641698
; GENERAL INFORMATION:
; APPLICANT: O'Malley, Bert W.
; APPLICANT: Tsai, Ming-Jer
; APPLICANT: Leдебур, Harry C.
; APPLICANT: Kittle, Joseph D.

```


| | | | |
|----|--------|--|--------|
| Db | 389857 | TTTTTAAGTGGGGTATTAATGACAAATAACTCAAATGATTAAGTGGAGATTTTAAATGACAAAT | 389798 |
| Qy | 378 | TCATTAGAAATATGTTCCCTCCCTCGTGGTTCAAATGTTTGTAAAGATTTGTTCTGTGTAAA | 437 |
| Db | 389797 | ACTGTATATTAATTATATCTCTCTCTAAATCTCAATTAATTTATTTATTTATTTGGGAAA | 389738 |
| Qy | 438 | TATG 441 | |
| Db | 389737 | TATG 389734 | |

RESULT 49
 US-08-545-528D-1/c
 ; Sequence 1, Application US/08545528D
 ; Patent No. 6537773
 ; GENERAL INFORMATION:
 ; APPLICANT: Fraser et al.
 ; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
 ; Patent No. 6537773
 ; TITLE OF INVENTION: Thereof, and Uses Thereof
 ; FILE REFERENCE: P6193P1
 ; CURRENT APPLICATION NUMBER: US/08/545,528D
 ; CURRENT FILING DATE: 1995-10-19
 ; PRIOR APPLICATION NUMBER: US 08/488,018
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: US 08/473,545
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 580073
 ; TYPE: DNA
 ; ORGANISM: Mycoplasma genitalium
 US-08-545-528D-1

| | Query Match | 7.3%; | Score 34.2; | DB 4; | Length 580073; |
|----|-----------------------|---|---------------|-------------------|-------------------|
| | Best Local Similarity | 52.4%; | Prod. No. 49; | | |
| | Matches | 75; | Conservative | 0; Mismatches 68; | Indels 0; Gaps 0; |
| Qy | 317 | TTATTAAAGTGTTTATGATCTTATACATAATTTATACAAAGATATTAAAGCCCTG | 376 | | |
| Db | 377708 | TTGGTGAAGTGTTATTATTTTGGTTTGTGTAGTACAGTTTTTCAAAATGTAAACCCCTT | | | |
| Qy | 377 | TTCAATTAAGAAATCTCCCTTCCCTGGTTCATGTTTGTAAAGATGTTCTGTGTAA | 436 | | |
| Db | 377648 | ATCACTAGCAAGATCAATTCGCTTGTGTTGATTCATTAAGAGGGGATGAAAAACA | | | |
| Qy | 437 | ATATGCTTTTATAAACAAGTT | 459 | | |
| Db | 377588 | AACTCAGTTTTAAATGTCACTT | 377566 | | |

```

RESULT 50
US-09-135-994-3/c
; Sequence 3, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-3
Query Match 7.3%; Score 34; DB 3; Length 300;

```

Search completed: May 25, 2004, 17:19:51
Job time : 101 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 17:19:57 ; Search time 92 Seconds
(without alignments)
2816.979 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED
Perfect score: 467
Sequence: 1 caatgttgctaccact.....taataaacagttaaaagctg 467

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 10

Total number of hits satisfying chosen parameters: 144238

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued Patents NA:
1: /cgm2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgm2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgm2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgm2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgm2_6/ptodata/2/ina/PCTUS COMB.seq:*
6: /cgm2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 467 | 100.0 | 50000 | 4 | US-09-146-053-4 |
| C 2 | 25 | 5.4 | 36 | 2 | US-08-863-639A-31 |
| C 3 | 25 | 5.4 | 91 | 4 | US-09-091-952A-15 |
| C 4 | 25 | 5.4 | 94 | 4 | US-09-091-952A-14 |
| C 5 | 25 | 5.4 | 397 | 3 | US-09-253-691-3 |
| C 6 | 25 | 5.4 | 432 | 4 | US-09-060-299-5 |
| C 7 | 25 | 5.4 | 432 | 4 | US-09-402-923A-5 |
| C 8 | 25 | 5.4 | 570 | 4 | US-09-621-976-91 |
| C 9 | 25 | 5.4 | 575 | 4 | US-09-621-976-2968 |
| C 10 | 25 | 5.4 | 841 | 4 | US-09-566-921-136 |
| C 11 | 25 | 5.4 | 1776 | 2 | US-08-531-927B-1 |
| C 12 | 25 | 5.4 | 1776 | 3 | US-09-041-886-12 |
| C 13 | 25 | 5.4 | 1884 | 3 | US-08-753-007A-5 |
| C 14 | 25 | 5.4 | 1884 | 3 | US-09-398-496-5 |
| C 15 | 25 | 5.4 | 2023 | 4 | US-09-491-522-6 |
| C 16 | 25 | 5.4 | 2128 | 4 | US-09-675-305-13 |
| C 17 | 25 | 5.4 | 2234 | 1 | US-08-261-822A-7 |
| C 18 | 25 | 5.4 | 2234 | 5 | PCT-US95-0774A-7 |
| C 19 | 25 | 5.4 | 2450 | 4 | US-09-491-522-2 |
| C 20 | 25 | 5.4 | 2770 | 3 | US-09-008-697A-13 |
| C 21 | 25 | 5.4 | 3231 | 4 | US-09-866-028-14 |
| C 22 | 25 | 5.4 | 4262 | 4 | US-09-521-511C-10 |
| C 23 | 25 | 5.4 | 4843 | 4 | US-09-060-299-2 |
| C 24 | 25 | 5.4 | 4843 | 4 | US-09-402-923A-2 |
| C 25 | 25 | 5.4 | 5098 | 4 | US-09-060-299-1 |
| C 26 | 25 | 5.4 | 5098 | 4 | US-09-402-923A-1 |
| C 27 | 25 | 5.4 | 6692 | 4 | US-09-491-522-1 |

ALIGNMENTS

```

RESULT 1
US-09-146-053-4/c
; Sequence 4, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-4

Query Match          100.0%; Score 467; DB 4; Length 50000;
Best Local Similarity 100.0%; Pred. No. 3.1e-223;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CAATGTTGCTCTATCCACTCCCGCAAGCCCTTACCTATGCTGCTGCTGCTGCTG 60
Db      1518 CAATGTTGCTCTATCCACTCCCGCAAGCCCTTACCTATGCTGCTGCTGCTGCTG 1459

QY      61  CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db      1458 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1399

QY      121 GTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Db      1398 GTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1339

QY      181 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db      1338 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279

QY      241 GAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 300
Db      1278 GAGAAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 1219

QY      301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATCTAATTATACAA 360
Db      1218 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATCTAATTATACAA 1159

QY      361 AGATATTAAAGCCCTGCTTCAATTAAGAAATGTTTCCTTCCCTGCTGTTCAATGTTGTAA 420
Db      1158 AGATATTAAAGCCCTGCTTCAATTAAGAAATGTTTCCTTCCCTGCTGTTCAATGTTGTAA 1099

QY      421 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTAAAGCTG 467
Db      1098 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTAAAGCTG 1052

```

RESULT 2

```

US-08-863-639A-31/c
; Sequence 31, Application US/08863639A
; Patent No. 5981185
; GENERAL INFORMATION:
; APPLICANT: Matson, Robert S.
; APPLICANT: Coassin, Peter J.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Caskey, C. T.

```

```

; TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Corel WordPerfect 8 version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,639A
; FILING DATE: May 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph E. Mueth
; REGISTRATION NUMBER: 20,532
; REFERENCE/DOCKET NUMBER: 11859-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
US-08-863-639A-31

Query Match          5.4%; Score 25; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      53  CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 77
Db      27  CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3

RESULT 3
US-09-091-952A-15
; Sequence 15, Application US/09091952A
; Patent No. 6458532
; GENERAL INFORMATION:
; APPLICANT: Detera-Wadleigh, Sevilla D.
; Gershon, Elliot S.
; Badner, Judith A.
; Goldin, Lynn R.
; Berrettini, Wade H.
; Yoshikawa, Takeo
; Sanders, Alan R.
; Esterling, Lisa E.
; TITLE OF INVENTION: Chromosomal Markers and Diagnostic
; Tests for Manic-Depressive Illness
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,952A

```

```

; FILING DATE: 19-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,278
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: PCT/US97/19381
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-297100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1...91
; OTHER INFORMATION: Clone 22 allele 2 polymorphic marker
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: 37...63
; OTHER INFORMATION: Clone 22 allele 2 polymorphic repeat
; sequence
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-091-952A-15

Query Match 5.4%; Score 25; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 40 GCTGCTGCTGCTGCTGCTGCTGCTT 64

RESULT 4
US-09-091-952A-14
; Sequence 14, Application US/09091952A
; Patent No. 6458532
; GENERAL INFORMATION:
; APPLICANT: Detera-Wadleigh, Sevilla D.
; Gershon, Elliot S.
; Badner, Judith A.
; Goldin, Lynn R.
; Berrettini, Wade H.
; Yoshikawa, Takeo
; Sanders, Alan R.
; Esterling, Lisa E.
; TITLE OF INVENTION: Chromosomal Markers and Diagnostic
; Tests for Manic-Depressive Illness
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,952A

```

```

; FILING DATE: 19-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,278
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: PCT/US97/19381
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-297100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1...94
; OTHER INFORMATION: Clone 22 allele 1 polymorphic marker
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: 37...66
; OTHER INFORMATION: Clone 22 allele 1 polymorphic repeat
; sequence
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-091-952A-14

Query Match 5.4%; Score 25; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 43 GCTGCTGCTGCTGCTGCTGCTGCTT 67

RESULT 5
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
; US-09-253-691-3

Query Match 5.4%; Score 25; DB 3; Length 397;
Best Local Similarity 100.0%; Pred. No. 0.0049;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCTT 77
Db 357 CGCTGCTGCTGCTGCTGCTGCTGCTT 333

```

Hey, Patricia
Kawaguchi, Yoshihiko
Merriman, Tony R
Metzker, Michael L
TITLE OF INVENTION: No. 6555654e1 LDL-Receptor
NUMBER OF SEQUENCES: 455
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon and Vanderhye
STREET: 1100 No. 6555654th Glebe Road, Eighth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: US
ZIP: VA 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/402,923A
FILING DATE: 14-Feb-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB98/01102
FILING DATE: 15-APR-1998
APPLICATION NUMBER: US 60/043,553
FILING DATE: 15-APR-1997
APPLICATION NUMBER: US 60/048,740
FILING DATE: 05-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: B.J.Sadoff
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 620-81
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)816-4091
TELEFAX: (703)816-4100
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-402-923A-5
Query Match 5.4%; Score 25; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 0.0049; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
DB 80 CGCTGCTGCTGCTGCTGCTGCTGCT 104
RESULT 8
US-09-621-976-91/c
Sequence 91, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 91
LENGTH: 570
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS

Hey, Patricia
Kawaguchi, Yoshihiko
Merriman, Tony R
Metzker, Michael L
TITLE OF INVENTION: No. 6555654e1 LDL-Receptor
NUMBER OF SEQUENCES: 455
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon and Vanderhye
STREET: 1100 No. 6555654th Glebe Road, Eighth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: US
ZIP: VA 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,299
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/043,553
FILING DATE: 15-APR-1997
APPLICATION NUMBER: US 60/048,740
FILING DATE: 05-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: B.J.Sadoff
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 620-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)816-4091
TELEFAX: (703)816-4100
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-060-299-5
Query Match 5.4%; Score 25; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 0.0049; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
DB 80 CGCTGCTGCTGCTGCTGCTGCTGCT 104
RESULT 7
US-09-402-923A-5
Sequence 5, Application US/09402923A
Patent No. 6555654
GENERAL INFORMATION:
APPLICANT: Todd, John A
APPLICANT: Hess, John W
APPLICANT: Caskey, Charles T
APPLICANT: Cox, Roger D
APPLICANT: Gerhold, David
APPLICANT: Hammond, Holly
TITLE OF INVENTION: No. 6545137e1 Receptor
NUMBER OF SEQUENCES: 455
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hey, Patricia
STREET: 1100 No. 6545137th Glebe Road, Eighth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: US
ZIP: VA 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,299
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/043,553
FILING DATE: 15-APR-1997
APPLICATION NUMBER: US 60/048,740
FILING DATE: 05-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: B.J.Sadoff
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 620-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)816-4091
TELEFAX: (703)816-4100
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-060-299-5
Query Match 5.4%; Score 25; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 0.0049; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
DB 80 CGCTGCTGCTGCTGCTGCTGCTGCT 104
RESULT 7
US-09-402-923A-5
Sequence 5, Application US/09402923A
Patent No. 6555654
GENERAL INFORMATION:
APPLICANT: Todd, John A
APPLICANT: Hess, John W
APPLICANT: Caskey, Charles T
APPLICANT: Cox, Roger D
APPLICANT: Gerhold, David
APPLICANT: Hammond, Holly

```
US-09-566-921-136
Query Match
Best Local Similarity 5.4%; Score 25; DB 4; Length 841;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 583 GCTGCTGCTGCTGCTGCTGCTGCTT 607

RESULT 11
US-08-531-927B-1/c
; Sequence 1, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531,927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATH95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1115
;
US-08-531-927B-1
Query Match
Best Local Similarity 5.4%; Score 25; DB 2; Length 1776;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGTGTGCTGCTGCTGCTGCTGCTT 77
Db 988 GCTGTGCTGCTGCTGCTGCTGCTT 964

RESULT 12
US-09-041-886-12/c
; Sequence 12, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
```



```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-261-822A-7

Query Match          5.4%; Score 25; DB 1; Length 2234;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db      1973 GCTGCTGCTGCTGCTGCTGCTGCTT 1949

RESULT 18
PCT-US95-07744A-7/c
; Sequence 7, Application PC/TUS9507744A
; GENERAL INFORMATION:
; APPLICANT: Trustees of The University of Pennsylvania
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norriis
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07744A
; FILING DATE: 15-JUNE-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,822
; FILING DATE: June 17, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-07744A-7

Query Match          5.4%; Score 25; DB 5; Length 2234;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db      1973 GCTGCTGCTGCTGCTGCTGCTGCTT 1949

RESULT 19
US-09-491-522-2
; Sequence 2, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-2

Query Match          5.4%; Score 25; DB 4; Length 2450;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
Db      125 CGCTGCTGCTGCTGCTGCTGCTGCT 149

RESULT 20
US-09-008-697A-13/c
; Sequence 13, Application US/09008697A
; Patent No. 6197504
; GENERAL INFORMATION:
; APPLICANT: Chow, King Lau
; TITLE OF INVENTION: USES OF MAB-21
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```



```
; APPLICATION NUMBER: US/09/008,697A
; FILING DATE: January 19, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 50752-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2770 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-008-697A-13

Query Match 5.4%; Score 25; DB 3; Length 2770;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTT 78
Db 619 GCTGCTGCTGCTGCTGCTGCTT 595

RESULT 21
US-09-866-028-14/c
; Sequence 14, Application US/09866028
; Patent No. 6642360
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Geritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kljavin, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/866,028
; CURRENT FILING DATE: 2001-05-25
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 14
; LENGTH: 3231
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-09-866-028-14

Query Match 5.4%; Score 25; DB 4; Length 3231;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCT 77
Db 459 CGCTGCTGCTGCTGCTGCTGCT 435

RESULT 22
US-09-521-511C-10/c
```

```
; Sequence 10, Application US/09521511C
; Patent No. 6555358
; GENERAL INFORMATION:
; APPLICANT: Gurnett, Anne
; APPLICANT: Liberator, Paul A.
; APPLICANT: Donald, Robert
; APPLICANT: Schmatz, Dennis
; APPLICANT: Harris, Georgianna
; APPLICANT: Rattray, Sandra J.
; TITLE OF INVENTION: CYCLIC GMP DEPENDENT PROTEIN KINASE AS A
; CHEMOTHERAPEUTIC TARGET FOR ANTIPROTOZOAL AGENTS
; FILE REFERENCE: 20362
; CURRENT APPLICATION NUMBER: US/09/521,511C
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/129,058
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 4262
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PKG
; US-09-521-511C-10

Query Match 5.4%; Score 25; DB 4; Length 4262;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGTGTGCTGCTGCTGCTGCTGCT 77
Db 2091 CGTGTGCTGCTGCTGCTGCTGCT 2067

RESULT 23
US-09-060-299-2
; Sequence 2, Application US/09060299
; Patent No. 6545137
; GENERAL INFORMATION:
; APPLICANT: Todd, John A
; APPLICANT: Hess, John W
; APPLICANT: Caskey, Charles T
; APPLICANT: Cox, Roger D
; APPLICANT: Gerhold, David
; APPLICANT: Hammond, Holly
; APPLICANT: Hey, Patricia
; APPLICANT: Kawaguchi, Yoshihiko
; APPLICANT: Merriman, Tony R
; APPLICANT: Metzker, Michael L
; TITLE OF INVENTION: No. 6545137el Receptor
; NUMBER OF SEQUENCES: 455
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon and Vanderhye
; STREET: 1100 No. 6545137th Glebe Road, Eighth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,299
; FILING DATE: 15-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/043,553
; FILING DATE: 15-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,740
; FILING DATE: 05-JUN-1997
```

```
;
;
; ATTORNEY/AGENT INFORMATION:
; NAME: B.J.Sadoff
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 620-35
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4091
; TELEFAX: (703)816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4843 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-060-299-2
Query Match 5.4%; Score 25; DB 4; Length 4843;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 32 CGCTGCTGCTGCTGCTGCTGCTGCT 56

RESULT 24
US-09-402-923A-2
; Sequence 2, Application US/09402923A
; Patent No. 6555654
; GENERAL INFORMATION:
; APPLICANT: Todd, John A
; APPLICANT: Hess, John W
; APPLICANT: Caskey, Charles T
; APPLICANT: Cox, Roger D
; APPLICANT: Gerhold, David
; APPLICANT: Hammond, Holly
; APPLICANT: Hey, Patricia
; APPLICANT: Kawaguchi, Yoshihiko
; APPLICANT: Merriman, Tony R
; APPLICANT: Metzker, Michael L
; TITLE OF INVENTION: No. 6555654el LDL-Receptor
; NUMBER OF SEQUENCES: 455
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon and Vandexhye
; STREET: 1100 No. 6555654th Glebe Road, Eighth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/402,923A
; FILING DATE: 14-Feb-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB98/01102
; FILING DATE: 15-APR-1998
; APPLICATION NUMBER: US 60/043,553
; FILING DATE: 15-APR-1997
; APPLICATION NUMBER: US 60/048,740
; FILING DATE: 05-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: B.J.Sadoff
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 620-81
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4091
; TELEFAX: (703)816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4843 base pairs
```

```
;
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-402-923A-2
Query Match 5.4%; Score 25; DB 4; Length 4843;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 32 CGCTGCTGCTGCTGCTGCTGCTGCT 56

RESULT 25
US-09-060-299-1
; Sequence 1, Application US/09060299
; Patent No. 6545137
; GENERAL INFORMATION:
; APPLICANT: Todd, John A
; APPLICANT: Hess, John W
; APPLICANT: Caskey, Charles T
; APPLICANT: Cox, Roger D
; APPLICANT: Gerhold, David
; APPLICANT: Hammond, Holly
; APPLICANT: Hey, Patricia
; APPLICANT: Kawaguchi, Yoshihiko
; APPLICANT: Merriman, Tony R
; APPLICANT: Metzker, Michael L
; TITLE OF INVENTION: No. 6545137el Receptor
; NUMBER OF SEQUENCES: 455
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon and Vandexhye
; STREET: 1100 No. 6545137th Glebe Road, Eighth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/060,299
; FILING DATE: 15-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/043,553
; FILING DATE: 15-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,740
; FILING DATE: 05-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: B.J.Sadoff
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 620-35
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4091
; TELEFAX: (703)816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5098 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-060-299-1
Query Match 5.4%; Score 25; DB 4; Length 5098;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
```

Db 80 CGCTGCTGCTGCTGCTGCTGCT 104

RESULT 26

US-09-402-923A-1

; Sequence 1, Application US/09402923A

; Patent No. 6555654

GENERAL INFORMATION:

APPLICANT: Todd, John A

; Hess, John W

; Caskey, Charles T

; Cox, Roger D

; Gerhold, David

; Hammond, Holly

; Hey, Patricia

; Kawaguchi, Yoshihiko

; Merriman, Tony R

; Metzker, Michael L

; TITLE OF INVENTION: No. 6555654el LDL-Receptor

; NUMBER OF SEQUENCES: 455

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon and Vanderhye

; STREET: 1100 No. 6555654th Glebe Road, Eighth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: US

; ZIP: VA 22201-4714

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/402,923A

; FILING DATE: 14-Feb-2001

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB98/01102

; FILING DATE: 15-APR-1998

; APPLICATION NUMBER: US 60/043,553

; FILING DATE: 15-APR-1997

; APPLICATION NUMBER: US 60/048,740

; FILING DATE: 05-JUN-1997

ATTORNEY/AGENT INFORMATION:

; NAME: B.J.Sadoff

; REGISTRATION NUMBER: 36,663

; REFERENCE/DOCKET NUMBER: 620-81

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)816-4091

; TELEFAX: (703)816-4100

SEQUENCE CHARACTERISTICS:

; LENGTH: 5098 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-402-923A-1

Query Match 5.4%; Score 25; DB 4; Length 5098;

Best Local Similarity 100.0%; Pred.No. 0.0052;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCT 77

Db 80 CGCTGCTGCTGCTGCTGCTGCT 104

RESULT 27

US-09-491-522-1

; Sequence 1, Application US/09491522

; Patent No. 6428998

GENERAL INFORMATION:

APPLICANT: Collige, Alain

; Lapiere, Charles M.

; Prockop, Darwin J.

; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE.

; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF

; NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds, LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036-2811

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/491,522

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/886,333

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Abrams, Samuel B

; REGISTRATION NUMBER: 30,605

; REFERENCE/DOCKET NUMBER: 8389-0060-999

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-493-4935

; TELEFAX: 650-493-5556

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6692 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-09-491-522-1

Query Match 5.4%; Score 25; DB 4; Length 6692;

Best Local Similarity 100.0%; Pred.No. 0.0052;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCT 77

Db 125 CGCTGCTGCTGCTGCTGCTGCT 149

RESULT 28

US-09-091-952A-6

; Sequence 6, Application US/09091952A

; Patent No. 6458532

GENERAL INFORMATION:

APPLICANT: Detera-Wadleigh, Sevilla D.

; Gershon, Elliot S.

; Badner, Judith A.

; Goldin, Lynn R.

; Berrettini, Wade H.

; Yoshikawa, Takeo

; Sanders, Alan R.

; Esterling, Lisa E.

; TITLE OF INVENTION: Chromosomal Markers and Diagnostic

; Tests for Manic-Depressive Illness

; NUMBER OF SEQUENCES: 197

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/091-952A
; FILING DATE: 19-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,278
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: PCT/US97/19381
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-297100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8065 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1...8065
; OTHER INFORMATION: Clone 22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 116...1036
; OTHER INFORMATION: Clone 22 coding region
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 452...505
; OTHER INFORMATION: alternatively spliced portion
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5595...5685
; OTHER INFORMATION: amplified region for genotyping
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-091-952A-6
Query Match 5.4%; Score 25; DB 4; Length 8065;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTT 78
Db 5634 GCTGCTGCTGCTGCTGCTGCTT 5658

RESULT 29
US-08-068-747-6/c
; Sequence 6, Application US/08068747
; Patent No. 5695933
; GENERAL INFORMATION:
; APPLICANT: Schalling, Martin
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: Direct Determination of Expanded
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts

```

```

;
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/068,747
; FILING DATE: 28-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic"
US-08-068-747-6
Query Match 5.1%; Score 24; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCT 5

RESULT 30
US-08-068-747-11
; Sequence 11, Application US/08068747
; Patent No. 5695933
; GENERAL INFORMATION:
; APPLICANT: Schalling, Martin
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: Direct Determination of Expanded
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/068,747
; FILING DATE: 28-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:

```

```
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic"
US-08-068-747-11

Query Match          5.1%; Score 24; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 3 GCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 31
US-08-863-639A-30/c
; Sequence 30, Application US/08863639A
; Patent No. 5981185
; GENERAL INFORMATION:
; APPLICANT: Matson, Robert S.
; APPLICANT: Coassin, Peter J.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Caskey, C. T.
; TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Corel WordPerfect 8 version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863.639A
; FILING DATE: May 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph E. Mueth
; REGISTRATION NUMBER: 20,532
; REFERENCE/DOCKET NUMBER: 11859-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
US-08-863-639A-30

Query Match          5.1%; Score 24; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5

RESULT 32
US-09-135-994-4/c
; Sequence 4, Application US/09135994A
; Patent No. 6280938

; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-4

Query Match          5.1%; Score 24; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5

RESULT 33
US-09-684-843A-4/c
; Sequence 4, Application US/09684843A
; Patent No. 6514755
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: Regents of the University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/684,843A
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/056,170
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 09/135,994
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-684-843A-4

Query Match          5.1%; Score 24; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5

RESULT 34
US-08-570-155-14/c
; Sequence 14, Application US/08570155
; Patent No. 5962332
; GENERAL INFORMATION:
; APPLICANT: Singer, Robert H.
; APPLICANT: Taneja, Krishan L.
; TITLE OF INVENTION: DETECTION OF TRINUCLEOTIDE REPEATS
; TITLE OF INVENTION: BY IN SITU HYBRIDIZATION
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,155
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/399,499
; FILING DATE: 07 March 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/214,823
; FILING DATE: 17 March 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06353/011001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; PCT-US95-02861-14
;
; Query Match 5.1%; Score 24; DB 2; Length 31;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
; Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5
;
; RESULT 35
; US-08-570-155-14
;
; Query Match 5.1%; Score 24; DB 2; Length 31;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
; Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5
;
; RESULT 36
; US-08-863-639A-29
; Sequence 29, Application US/08863639A
; Patent No. 5981185
; GENERAL INFORMATION:
; APPLICANT: Matson, Robert S.
; APPLICANT: Coassin, Peter J.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Caskey, C. T.
; TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Corel WordPerfect 8 version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,639A
; FILING DATE: May 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph E. Mueth
; REGISTRATION NUMBER: 20,532
; REFERENCE/DOCKET NUMBER: 11859-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; US-08-863-639A-29
;
; Query Match 5.1%; Score 24; DB 2; Length 33;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; PCT-US95-02861-14/c
; Sequence 14, Application PC/TUS9502861
; GENERAL INFORMATION:
; APPLICANT: Singer, Robert H.
; APPLICANT: Taneja, Krishan L.
; TITLE OF INVENTION: DETECTION OF TRINUCLEOTIDE
; TITLE OF INVENTION: REPEATS
; TITLE OF INVENTION: BY IN SITU HYBRIDIZATION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version
; SOFTWARE: #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02861
; FILING DATE: 08 March 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/214,823
; FILING DATE: 17 March 1994
```

```
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 3 GCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 37
US-08-068-747-1
; Sequence 1, Application US/08068747
; Patent No. 5695933
; GENERAL INFORMATION:
; APPLICANT: Schalling, Martin
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: Direct Determination of Expanded
; TITLE OF INVENTION: Nucleotide Repeats in the Human Genome
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/068,747
; FILING DATE: 28-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-068-747-1

Query Match 5.1%; Score 24; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 3 GCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 38
US-09-136-080E-50
; Sequence 50, Application US/09136080E
; Patent No. 6518017
; GENERAL INFORMATION:
; APPLICANT: Riley, Timothy A.
; APPLICANT: Brown, Bob D.
; APPLICANT: Arnold, Lyle J.
; TITLE OF INVENTION: COMBINATORIAL ANTISENSE LIBRARY
; FILE REFERENCE: OASHIO.003A
; CURRENT APPLICATION NUMBER: US/09/136,080E
; CURRENT FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
```

```
; LENGTH: 53
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic oligonucleotide
US-09-136-080E-50

Query Match 5.1%; Score 24; DB 4; Length 53;
Best Local Similarity 66.7%; Pred. No. 0.014;
Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 30 GCUGCUGCUGCUGCUGCUGCUGCU 53

RESULT 39
US-09-043-303-14/c
; Sequence 14, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJII, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 57
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-14

Query Match 5.1%; Score 24; DB 3; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 46 GCTGCTGCTGCTGCTGCTGCTGCT 23

RESULT 40
US-09-043-303-13/c
; Sequence 13, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJII, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 69
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-13

Query Match 5.1%; Score 24; DB 3; Length 69;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
 Db 40 GCTGCTGCTGCTGCTGCTGCT 17

RESULT 41

US-09-043-303-10/c
 ; Sequence 10, Application US/09043303
 ; Patent No. 6251589
 ; GENERAL INFORMATION:
 ; APPLICANT: TSUJI, Shoji
 ; APPLICANT: SANPEI, Kazujiro
 ; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
 ; FILE REFERENCE: 0760-0241P
 ; CURRENT APPLICATION NUMBER: US/09/043,303
 ; CURRENT FILING DATE: 1998-05-18
 ; EARLIER APPLICATION NUMBER: PCT/JP96/01999
 ; EARLIER FILING DATE: 1996-07-18
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 75
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-043-303-10

Query Match 5.1%; Score 24; DB 3; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
 Db 25 GCTGCTGCTGCTGCTGCTGCT 2

RESULT 42

US-09-043-303-11/c
 ; Sequence 11, Application US/09043303
 ; Patent No. 6251589
 ; GENERAL INFORMATION:
 ; APPLICANT: TSUJI, Shoji
 ; APPLICANT: SANPEI, Kazujiro
 ; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
 ; FILE REFERENCE: 0760-0241P
 ; CURRENT APPLICATION NUMBER: US/09/043,303
 ; CURRENT FILING DATE: 1998-05-18
 ; EARLIER APPLICATION NUMBER: PCT/JP96/01999
 ; EARLIER FILING DATE: 1996-07-18
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 11
 ; LENGTH: 78
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-043-303-11

Query Match 5.1%; Score 24; DB 3; Length 78;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
 Db 67 GCTGCTGCTGCTGCTGCTGCT 44

RESULT 43

US-09-043-303-12/c
 ; Sequence 12, Application US/09043303
 ; Patent No. 6251589
 ; GENERAL INFORMATION:
 ; APPLICANT: TSUJI, Shoji

; APPLICANT: SANPEI, Kazujiro
 ; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
 ; FILE REFERENCE: 0760-0241P
 ; CURRENT APPLICATION NUMBER: US/09/043,303
 ; CURRENT FILING DATE: 1998-05-18
 ; EARLIER APPLICATION NUMBER: PCT/JP96/01999
 ; EARLIER FILING DATE: 1996-07-18
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 12
 ; LENGTH: 78
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-043-303-12

Query Match 5.1%; Score 24; DB 3; Length 78;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
 Db 67 GCTGCTGCTGCTGCTGCTGCT 44

RESULT 44

US-09-205-995-75/c
 ; Sequence 75, Application US/09205995
 ; Patent No. 6368855
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Minzhen
 ; APPLICANT: Qiu, Gang
 ; APPLICANT: Humphreys, Robert
 ; TITLE OF INVENTION: CANCER CELL VACCINE
 ; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
 ; CURRENT APPLICATION NUMBER: US/09/205,995
 ; CURRENT FILING DATE: 1998-12-04
 ; PRIOR APPLICATION NUMBER: 09/036,746
 ; PRIOR FILING DATE: 1998-03-09
 ; PRIOR APPLICATION NUMBER: 08/661,627
 ; PRIOR FILING DATE: 1996-06-11
 ; NUMBER OF SEQ ID NOS: 79
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 75
 ; LENGTH: 107
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Reverse gene
 ; OTHER INFORMATION: construct corresponding to a specific region of
 ; OTHER INFORMATION: the mouse li gene.
 US-09-205-995-75

Query Match 5.1%; Score 24; DB 4; Length 107;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
 Db 42 GCTGCTGCTGCTGCTGCTGCT 19

RESULT 45

US-08-469-802B-6/c
 ; Sequence 6, Application US/08469802B
 ; Patent No. 5741645
 ; GENERAL INFORMATION:
 ; APPLICANT: Orr, Harry T.
 ; APPLICANT: Ratum, Laura P.W.
 ; APPLICANT: Chung, Ming-yi
 ; APPLICANT: Zoghbi, Huda Y.
 ; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
 ; Patent No. 5741645

```
/ TITLE OF INVENTION: Type 1 and Method for Diagnosis
/ NUMBER OF SEQUENCES: 47
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Mueiting, Raasch, Gebhardt & Schwappach, P.A.
/ STREET: 119 No. 5741645th Fourth Street, Suite 203
/ CITY: Minneapolis
/ STATE: MN
/ COUNTRY: USA
/ ZIP: 55401
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/469,802B
/ FILING DATE: 06-JUN-1995
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Mueiting, Ann M.
/ REGISTRATION NUMBER: 33,977
/ REFERENCE/DOCKET NUMBER: 110.00030101
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 612-305-1217
/ TELEFAX: 612-305-1225
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 154 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
/ US-08-469-802B-6

Query Match 5.1%; Score 24; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCT 77
DB 152 GCTGCTGCTGCTGCTGCTGCT 129

RESULT 45
US-08-267-803B-6/c
/ Sequence 6, Application US/08267803B
/ Patent No. 5834183
/ GENERAL INFORMATION:
/ APPLICANT: Orr, Harty T.
/ APPLICANT: Ranum, Laura P.W.
/ APPLICANT: Chung, Ming-yi
/ APPLICANT: Zoghbi, Huda Y.
/ TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
/ Patent No. 5834183
/ TITLE OF INVENTION: Type 1 and Method for Diagnosis
/ NUMBER OF SEQUENCES: 85
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Mueiting, Raasch, Gebhardt & Schwappach, P.A.
/ STREET: P.O. Box 581415
/ CITY: Minneapolis
/ STATE: MN
/ COUNTRY: USA
/ ZIP: 55458-1415
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/267,803B
/ FILING DATE: 28-JUN-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
```

```
/ NAME: McCormack, Myra H.
/ REGISTRATION NUMBER: 36,602
/ REFERENCE/DOCKET NUMBER: 110.00030120
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 612-305-1217
/ TELEFAX: 612-305-1228
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 154 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
/ US-08-267-803B-6

Query Match 5.1%; Score 24; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCT 77
DB 152 GCTGCTGCTGCTGCTGCTGCT 129

RESULT 47
US-08-332-766A-13/c
/ Sequence 13, Application US/08332766A
/ Patent No. 5843647
/ GENERAL INFORMATION:
/ APPLICANT: JEFFREYS, Alec J.
/ APPLICANT: ARMOUR, John
/ TITLE OF INVENTION: SIMPLE TANDEM REPEATS
/ NUMBER OF SEQUENCES: 125
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
/ STREET: 1100 New York Avenue, N.W.
/ CITY: Washington
/ STATE: D. C.
/ COUNTRY: U.S.A.
/ ZIP: 20005-3918
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/332,766A
/ FILING DATE: 01-NOV-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: GB 9326052.9
/ FILING DATE: 21-DEC-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIRD, Donald J.
/ REGISTRATION NUMBER: 25,323
/ REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 861-3000
/ TELEFAX: (202) 822-0944
/ TELEX: 6714627 CUSH
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 160 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-332-766A-13

Query Match 5.1%; Score 24; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 105 GCTGCTGCTGCTGCTGCTGCTGCT 82

RESULT 48
US-09-043-303-17/c
; Sequence 17, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJII, Shoji
; APPLICANT: SANPEI, Kazuhiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 165
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Probe
US-09-043-303-17

Query Match 5.1%; Score 24; DB 3; Length 165;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 163 GCTGCTGCTGCTGCTGCTGCTGCT 140

RESULT 49
US-08-469-802B-4/c
; Sequence 4, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueeting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
```

```
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-4

Query Match 5.1%; Score 24; DB 1; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 142 GCTGCTGCTGCTGCTGCTGCTGCT 119

RESULT 50
US-08-267-803B-4/c
; Sequence 4, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-267-803B-4

Query Match 5.1%; Score 24; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 142 GCTGCTGCTGCTGCTGCTGCTGCT 119
```

Search completed: May 25, 2004, 19:20:24
Job time : 94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 16:30:32 ; Search time 3345 Seconds
(without alignments)
6051.180 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctatccacct.....taataaacagttaaagctg 467

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 10

Total number of hits satisfying chosen parameters: 1846324

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|------------|--------------------|
| 1 | 467 | 100.0 | 494 | 6 | AX092298 | Sequence |
| 2 | 467 | 100.0 | 494 | 6 | AX696953 | Sequence |
| 3 | 467 | 100.0 | 495 | 9 | AY358657 | Homo sapi |
| 4 | 467 | 100.0 | 2673 | 9 | BC021104 | Homo sapi |
| 5 | 467 | 100.0 | 50000 | 6 | AR211704 | Sequence |
| 6 | 467 | 100.0 | 151152 | 9 | HS454M7 | Human DNA |
| 7 | 467 | 100.0 | 206618 | 9 | AF195953 | Homo sapi |
| 8 | 413 | 88.4 | 2243 | 6 | AX877684 | Sequence |
| 9 | 413 | 88.4 | 2243 | 6 | BD156791 | Primer fo |
| 10 | 413 | 88.4 | 2243 | 9 | AX001855 | Homo sapi |
| 11 | 241 | 51.6 | 458 | 6 | AX332625 | Sequence |
| 12 | 241 | 51.6 | 458 | 6 | AX332852 | Sequence |
| 13 | 238 | 51.0 | 403 | 6 | AX874329 | Sequence |
| 14 | 238 | 51.0 | 403 | 6 | BD154391 | Primer fo |
| 15 | 127 | 27.2 | 598 | 9 | HUMED38F03 | Homo sapi |
| 16 | 28 | 6.0 | 202198 | 2 | AC122549 | Mus muscu |
| 17 | 28 | 6.0 | 204171 | 2 | AC131317 | Mus muscu |
| 18 | 27 | 5.8 | 227 | 8 | AY023222 | Oryza sat |
| 19 | 27 | 5.8 | 373 | 11 | G09506 | human STS C |
| 20 | 27 | 5.8 | 1264 | 8 | AB028183 | Oryza sat |
| 21 | 27 | 5.8 | 1283 | 8 | AX073848 | Oryza sat |
| 22 | 27 | 5.8 | 2000 | 6 | AX654994 | Sequence |
| 23 | 27 | 5.8 | 75317 | 8 | SC4357 | Z38059 S.cerevisia |
| 24 | 27 | 5.8 | 88088 | 8 | AP004331 | Oryza sat |
| 25 | 27 | 5.8 | 128403 | 2 | AC017472 | Oryza sat |
| 26 | 27 | 5.8 | 132059 | 8 | AP003411 | Oryza sat |
| 27 | 27 | 5.8 | 143558 | 2 | AC103357 | Mus muscu |
| 28 | 27 | 5.8 | 158606 | 2 | AC138056 | Medicago |
| 29 | 27 | 5.8 | 163466 | 3 | AC010921 | Drosophil |
| 30 | 27 | 5.8 | 170178 | 5 | AL954848 | Zebrafish |
| 31 | 27 | 5.8 | 172069 | 3 | AC012160 | Drosophil |
| 32 | 27 | 5.8 | 172332 | 2 | AC137828 | Medicago |
| 33 | 27 | 5.8 | 177062 | 10 | AC121924 | Mus muscu |
| 34 | 27 | 5.8 | 185722 | 2 | AC102553 | Mus muscu |
| 35 | 27 | 5.8 | 188834 | 10 | AC098642 | Genomic s |
| 36 | 27 | 5.8 | 190402 | 2 | AC114715 | Rattus no |
| 37 | 27 | 5.8 | 192539 | 2 | AC132877 | Mus muscu |
| 38 | 27 | 5.8 | 192629 | 2 | AC118589 | Mus muscu |
| 39 | 27 | 5.8 | 194209 | 10 | AL929249 | Mouse DNA |
| 40 | 27 | 5.8 | 195835 | 10 | AL845432 | Mouse DNA |
| 41 | 27 | 5.8 | 210174 | 10 | AC124555 | Mus muscu |
| 42 | 27 | 5.8 | 211680 | 10 | AL603682 | Mouse DNA |
| 43 | 27 | 5.8 | 217278 | 10 | AL604024 | Mouse DNA |
| 44 | 27 | 5.8 | 237094 | 2 | AC116661 | Mus muscu |
| 45 | 27 | 5.8 | 237332 | 2 | AC131846 | Rattus no |
| 46 | 27 | 5.8 | 240501 | 2 | AC099207 | Rattus no |
| 47 | 27 | 5.8 | 240561 | 2 | AC129456 | Rattus no |
| 48 | 27 | 5.8 | 242978 | 2 | AC093480 | Mus muscu |
| 49 | 27 | 5.8 | 250546 | 2 | AC125723 | Rattus no |
| 50 | 27 | 5.8 | 258061 | 2 | AC128474 | Rattus no |
| 51 | 27 | 5.8 | 263257 | 2 | AC125564 | Rattus no |
| 52 | 27 | 5.8 | 264186 | 2 | AC131577 | Mus muscu |
| 53 | 27 | 5.8 | 298641 | 3 | AE003504 | Drosophil |
| 54 | 26 | 5.6 | 259 | 10 | AF187894 | AF187894 Dipodomys |
| 55 | 26 | 5.6 | 270 | 3 | AY333192 | Anopheles |
| 56 | 26 | 5.6 | 339 | 3 | AY333202 | Anopheles |
| 57 | 26 | 5.6 | 376 | 11 | G09495 | human STS C |
| 58 | 26 | 5.6 | 381 | 3 | AY333201 | Anopheles |
| 59 | 26 | 5.6 | 393 | 3 | AY333203 | Anopheles |
| 60 | 26 | 5.6 | 393 | 3 | AY333204 | Anopheles |
| 61 | 26 | 5.6 | 412 | 3 | AY333205 | Anopheles |
| 62 | 26 | 5.6 | 423 | 3 | AY333206 | Anopheles |
| 63 | 26 | 5.6 | 457 | 8 | FVE508245 | Fragaria |
| 64 | 26 | 5.6 | 502 | 8 | AY158013 | Fuacaeae |
| 65 | 26 | 5.6 | 735 | 10 | BC048870 | Mus muscu |

Pred. No. is the number of results predicted by chance to have a

```

66 5.6 1022 10 MMU080893 U08093 Mus musculus
67 5.6 1048 10 BC012681 BC012681 Mus muscu
68 5.6 1156 10 MMU080889 U08089 Mus muscu
69 5.6 1232 3 AY118433 AY118433 Drosophi
70 5.6 1685 10 BC036332 BC036332 Mus muscu
71 5.6 3063 3 AY060995 AY060995 Drosophi
72 5.6 7072 3 AB090813 AB090813 Anophele
73 5.6 7183 2 AC012983 AC012983 Drosophi
74 5.6 7673 10 AF026032 AF026032 Mus muscu
75 5.6 14532 2 AC019984 AC019984 Drosophi
76 5.6 15911 3 AF096709 AF096709 Drosophi
77 5.6 25724 2 AC019732 AC019732 Drosophi
78 5.6 36328 3 LMFL5213 LMFL5213 Leishmani
79 5.6 37891 2 AC014982 AC014982 Drosophi
80 5.6 38418 3 AY190934 AY190934 Drosophi
81 5.6 45300 8 AP005115 AP005115 Oryza sat
82 5.6 45375 5 AF328738 AF328738 Agelaliu
83 5.6 46427 3 AY190956 AY190956 Drosophi
84 5.6 46904 2 AC100911 AC100911 Mus muscu
85 5.6 56366 2 AC019800 AC019800 Drosophi
86 5.6 65464 2 AC121129 AC121129 Mus muscu
87 5.6 78456 10 AL627393 AL627393 Mouse DNA
88 5.6 78964 2 AC020460 AC020460 Drosophi
89 5.6 79786 6 AX695800 AX695800 Sequence
90 5.6 95946 2 AC138274 AC138274 Mus muscu
91 5.6 100500 2 AC133924 AC133924 Magnaport
92 5.6 110000 2 AC110108_1 Continuation (2 of
93 5.6 110000 2 AC110108_2 Continuation (3 of
94 5.6 110000 2 AC110642_2 Continuation (3 of
95 5.6 110000 2 AC110959_0 AC110959 Rattus no
96 5.6 110000 2 AC116234_1 Continuation (2 of
97 5.6 110000 2 LMFLCHR31_18 Continuation (19 o
98 5.6 110325 8 AP005003 AP005003 Oryza sat
99 5.6 113887 10 BX539328 BX539328 Mouse DNA
100 5.6 114135 2 AC137652 AC137652 Bos tauru

```

ALIGNMENTS

```

RESULT 1
AX092298 494 bp DNA linear PAT 21-MAR-2001
LOCUS
DEFINITION Sequence 29 from Patent WO0116318.
ACCESSION AX092298
VERSION AX092298.1 GI:134444463
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ferrera, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,
Eaton, D.L., Gao, W.Q., Fan, J., Hotstein, D., Fong, S., Goddard, A.,
Godowski, P.J., Gurney, A.L., Smith, V., Tumas, D., Wood, W.I.,
Grimaldi, C.J., Hillan, K.J., Paoni, N.P., Roy, M.A. and Watanabe, C.K.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 29 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source
Location/Qualifiers
1..494
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATGTTTGCCCTATCCACTCCCGCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
Db |
1 CAATGTTTGCCCTATCCACTCCCGCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGAGACTGCTCGGTGCCAGAAA 120
Db |
61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGAGACTGCTCGGTGCCAGAAA 120
QY 121 GTCTCTTCTGCCACTGAACGCCCTCCATCAGGAGATTGGGCCCTTCTTCCCTCCCTTCTCTTCTG 180
Db |
121 GTCTCTTCTGCCACTGAACGCCCTCCATCAGGAGATTGGGCCCTTCTTCCCTCCCTTCTCTTCTG 180
QY 181 TGTCTCTGCTCATCGGCTGCGCTGACCTGAGCTGAGCCCAAGCCCGCCGCTGGGAGGG 240
Db |
181 TGTCTCTGCTCATCGGCTGCGCTGAGCTGAGCCCAAGCCCGCCGCTGGGAGGG 240
QY 241 GAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAAAGAGGGTAGTGGGTG 300
Db |
241 GAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAAAGAGGGTAGTGGGTG 300
QY 301 GGTCTAGGGGGCTGCTCTTATTAAGTGTGTTTATGATTTCTTATCTAATTAACAA 360
Db |
301 GGTCTAGGGGGCTGCTCTTATTAAGTGTGTTTATGATTTCTTATCTAATTAACAA 360
QY 361 AGATATTAAAGGCCCTGTTTCATTAAGAAATTTTCCCTCCCTGCTGTTCAATGTTTGTAA 420
Db |
361 AGATATTAAAGGCCCTGTTTCATTAAGAAATTTTCCCTCCCTGCTGTTCAATGTTTGTAA 420
QY 421 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTTAAAGCTG 467
Db |
421 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTTAAAGCTG 467

```

RESULT 2

```

AX696953 494 bp DNA linear PAT 02-APR-2003
LOCUS
DEFINITION Sequence 21 from Patent WO0078961.
ACCESSION AX696953
VERSION AX696953.1 GI:29497962
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ferrera, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,
Eaton, D.L., Gao, W.Q., Fan, J., Hotstein, D., Fong, S., Goddard, A.,
Godowski, P.J., Gurney, A.L., Smith, V., Tumas, D., Wood, W.I.,
Grimaldi, C.J., Hillan, K.J., Paoni, N.P., Roy, M.A. and Watanabe, C.K.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0078961-A 21 28-DEC-2000;
Genentech, Inc. (US)
FEATURES
source
Location/Qualifiers
1..494
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATGTTTGCCCTATCCACTCCCGCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
Db |
1 CAATGTTTGCCCTATCCACTCCCGCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 181 TGTCTCTGCTCATCGGCTGCGATGACCTGCGAGCCAGCCAGCCCGCGGAGGG 240
Db 181 TGTCTCTGCTCATCGGCTGCGATGACCTGCGAGCCAGCCAGCCCGCGGAGGG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGGAGTGGGTG 300
Db 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGGAGTGGGTG 300
QY 301 GGCTAGGGGGCTGCTTATTATAAGTGGTGTATGATCTTATATCTTATATCAAA 360
Db 301 GGCTAGGGGGCTGCTTATTATAAGTGGTGTATGATCTTATATCTTATATCAAA 360
QY 361 AGATATTAAAGCCCTGTTTATTAAGTGGTGTATGATCTTATATCTTATATCAAA 420
Db 361 AGATATTAAAGCCCTGTTTATTAAGTGGTGTATGATCTTATATCTTATATCAAA 420
QY 421 AGATTTGTTCTGTGTAATATGCTTTTATAAAGTGGTGTATGATCTTATATCAAA 467
Db 421 AGATTTGTTCTGTGTAATATGCTTTTATAAAGTGGTGTATGATCTTATATCAAA 467

RESULT 3
AY358657
LOCUS
DEFINITION Homo sapiens clone DNA56862 APELIN (UNQ471) mRNA, complete cds.
ACCESSION AY358657
VERSION AY358657.1 GI:37182435
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Clark H.F., Gurney A.L., Abaya, E., Baker, K., Baldwin, D., Brush, J.,
Chen, J., Chow, B., Chui, C., Crowley, C., Currell, B., Deuel, B.,
Dowd, J., Eaton, D., Foster, J., Grimaldi, C., Gu, Q., Hass, P.E.,
Heldens, S., Huang, A., Kim, H.S., Klimowski, L., Jin, Y., Johnson, S.,
Lee, J., Lewis, L., Liao, D., Mark, M., Robbie, E., Sanchez, C.,
Schoenfeld, J., Seshagiri, S., Simmons, L., Singh, J., Smith, V.,
Stinson, J., Vagts, A., Vanden, R., Watanabe, C., Wietand, D., Woods, K.,
Xie, M.H., Yansura, D., Yi, S., Yu, G., Yuan, J., Zhang, M., Zhang, Z.,
Goddard, A., Wood, W.I. and Godowski, P.
The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
Effort to Identify Novel Human Secreted and Transmembrane Proteins:
A Bioinformatics Assessment
Genome Res. 13 (10), 2265-2270 (2003)
12975309
2. (bases 1 to 495)
Clark, H.F.
Direct Submission
Submitted (01-AUG-2003) Department of Bioinformatics, Genentech,
Inc., 1 DNA Way, South San Francisco, CA 94080, USA
FEATURES
source
1. .495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="DNA56862"
1. .495
/locus_tag="UNQ471"
40. .261
/locus_tag="UNQ471"
/notes="PRO831"
/codon_start=1
/product="APELIN"
/protein_id="AAQ89020.1"
/db_xref="GI:37182436"
/translations="MLLTLLILLKLGSGEWLVGAKVSSATDPIRDWAFPP
SFLLCLPHRPAMTCSQAQPRGEKVGDG"

ORIGIN
Query Match 100.0%; Score 467; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;


```

```

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTTGGCTATPCACCTCCCAAGCCCTTTACCTATGCTGCTACTAACGCTGCTG 60
Db 1 CAATGTTTGGCTATPCACCTCCCAAGCCCTTTACCTATGCTGCTACTAACGCTGCTG 60
QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTGAGTGGGACTGCTGCTGCTGCTGCTGCTG 120
Db 61 CTGCTGCTGCTGCTTAAAGCTCATGCTGAGTGGGACTGCTGCTGCTGCTGCTGCTG 120
QY 121 GTCTCTTCTCCCACTAGCGCCCTCATCAGGATTTGGGCTTCTTTCCCTCCCTTCTTCTG 180
Db 121 GTCTCTTCTCCCACTAGCGCCCTCATCAGGATTTGGGCTTCTTTCCCTCCCTTCTTCTG 180
QY 181 TGCTCTCTGCTCATCGGCTGCTGATGACCTGCGAGCAAGCCCGCTGGGGAGGG 240
Db 181 TGCTCTCTGCTCATCGGCTGCTGATGACCTGCGAGCAAGCCCGCTGGGGAGGG 240
QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGGAGTGGGTG 300
Db 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGGAGTGGGTG 300
QY 301 GGCTAGGGGGCTGCTTATTATAAGTGGTGTATGATCTTATATCTTATATCAAA 360
Db 301 GGCTAGGGGGCTGCTTATTATAAGTGGTGTATGATCTTATATCTTATATCAAA 360
QY 361 AGATATTAAAGCCCTGTTTATTAAGTGGTGTATGATCTTATATCTTATATCAAA 420
Db 361 AGATATTAAAGCCCTGTTTATTAAGTGGTGTATGATCTTATATCTTATATCAAA 420
QY 421 AGATTTGTTCTGTGTAATATGCTTTTATAAAGTGGTGTATGATCTTATATCAAA 467
Db 421 AGATTTGTTCTGTGTAATATGCTTTTATAAAGTGGTGTATGATCTTATATCAAA 467

BC021104 2673 bp mRNA linear PRI 07-OCT-2003
Homo sapiens apelin, mRNA (cDNA clone MGC:31846 IMAGE:4586949),
complete cds.
BC021104
BC021104.1 GI:18088893
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22386257
12477932
REFERENCE
2 (bases 1 to 2673)
Straussberg, R.
AUTHORS

```


| | | |
|---------------------------|--|--|
| repeat_region | /note="MER33 repeat: matches 1. .323 of consensus" | |
| repeat_region | 7178. .8315 | |
| repeat_region | /note="L1P2 repeat: matches 5000. .6146 of consensus" | |
| repeat_region | 11245. .11294 | |
| repeat_region | /note="MER5B repeat: matches 109. .178 of consensus" | |
| repeat_region | 11345. .11657 | |
| repeat_region | /note="AluYb8 repeat: matches 1. .310 of consensus" | |
| repeat_region | 11736. .11797 | |
| repeat_region | /note="31 copies 2 mer tt 72% conserved" | |
| repeat_region | 12436. .12562 | |
| repeat_region | /note="L2 repeat: matches 2579. .2710 of consensus" | |
| repeat_region | 13276. .13457 | |
| repeat_region | /note="AluSg/x repeat: matches 129. .310 of consensus" | |
| repeat_region | 13679. .13873 | |
| repeat_region | /note="MIR repeat: matches 5. .213 of consensus" | |
| repeat_region | 14132. .14206 | |
| repeat_region | /note="MIR repeat: matches 163. .233 of consensus" | |
| repeat_region | 14207. .14413 | |
| repeat_region | /note="MER8 repeat: matches 2. .239 of consensus" | |
| repeat_region | 14414. .14548 | |
| repeat_region | /note="MIR repeat: matches 12. .163 of consensus" | |
| repeat_region | 14793. .15073 | |
| repeat_region | /note="AluY repeat: matches 1. .311 of consensus" | |
| repeat_region | 15383. .15589 | |
| repeat_region | /note="MER3 repeat: matches 1. .209 of consensus" | |
| repeat_region | 16139. .16279 | |
| repeat_region | /note="MER5B repeat: matches 29. .178 of consensus" | |
| repeat_region | 16317. .16466 | |
| repeat_region | /note="MIR repeat: matches 52. .212 of consensus" | |
| repeat_region | 16667. .17000 | |
| repeat_region | /note="L1P9 repeat: matches 5829. .6163 of consensus" | |
| repeat_region | 17001. .17301 | |
| repeat_region | /note="AluYb8 repeat: matches 1. .302 of consensus" | |
| repeat_region | 17302. .17642 | |
| repeat_region | /note="L1P9 repeat: matches 5491. .5829 of consensus" | |
| repeat_region | 18285. .18620 | |
| repeat_region | /note="2 copies 168 mer 78% conserved" | |
| repeat_region | 18772. .18981 | |
| repeat_region | /note="MIR repeat: matches 6. .242 of consensus" | |
| repeat_region | 18945. .19005 | |
| repeat_region | /note="L2 repeat: matches 2648. .2702 of consensus" | |
| repeat_region | 20349. .20489 | |
| repeat_region | /note="MIR repeat: matches 1. .150 of consensus" | |
| repeat_region | 20543. .20697 | |
| repeat_region | /note="MER5A repeat: matches 22. .188 of consensus" | |
| repeat_region | 21496. .21830 | |
| repeat_region | /note="L1MC4 repeat: matches 7477. .7849 of consensus" | |
| repeat_region | 22420. .22650 | |
| repeat_region | /note="MER46A repeat: matches 1. .236 of consensus" | |
| repeat_region | 22894. .22965 | |
| repeat_region | /note="3 copies 24 mer 83% conserved" | |
| repeat_region | 24309. .24606 | |
| repeat_region | /note="AluX repeat: matches 1. .299 of consensus" | |
| repeat_region | 25081. .25134 | |
| repeat_region | /note="27 copies 2 mer ta 70% conserved" | |
| repeat_region | 25168. .25193 | |
| repeat_region | /note="13 copies 2 mer tg 100% conserved" | |
| Query Match | 100.0%; Score 467; DB 9; Length 151152; | |
| Best Local Similarity | 100.0%; Pred. No. 2.7e-258; | |
| Matches 467; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | |
| Qy | 1 CAATGTTGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTTAACGCTGCTG 60 | |
| Db | 89718 CAATGTTGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTTAACGCTGCTG 89659 | |
| Qy | 61 CTGCTGCTGCTGCTTAAGCTCATGCTTGAGTGGGACTGTCGTCGCCAGAAA 120 | |
| Db | 89658 CTGCTGCTGCTGCTTAAGGCTCATGCTTGAGTGGGACTGTCGTCGCCAGAAA 89599 | |
| Qy | 121 GTCTCTTCTGCCACTGACGCCCATCAGGAGTTGGCCCTTCTTCCCTTCCTTCTG 180 | |
| Db | 89598 GTCTCTTCTGCCACTGACGCCCATCAGGAGTTGGCCCTTCTTCCCTTCCTTCTG 89539 | |

| | | |
|------------|--|------------------------|
| Qy | 181 TGTCTCTGCTCATCGGCTGCGCATGACCTGCGAGCCAGCCCGCTGGGAAGGG 240 | |
| Db | 89538 TGTCTCTGCTCATCGGCTGCGCATGACCTGCGAGCCAGCCCGCTGGGAAGGG 89479 | |
| Qy | 241 GAGAAAGTGGGGATGCTAAGAAAGCTGGAGATAGGGAACAGAGGGTACTGGGTG 300 | |
| Db | 89478 GAGAAAGTGGGGATGCTAAGAAAGCTGGAGATAGGGAACAGAGGGTACTGGGTG 89419 | |
| Qy | 301 GGCTAGGGGGCTGCCCTTATTAAAGTGGTGTGTTATGATCTTATCTAATTTATACAA 360 | |
| Db | 89418 GGCTAGGGGGCTGCCCTTATTAAAGTGGTGTGTTATGATCTTATCTAATTTATACAA 89359 | |
| Qy | 361 AGATATTAAAGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGTTCAATGTTGTAA 420 | |
| Db | 89358 AGATATTAAAGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGTTCAATGTTGTAA 89299 | |
| Qy | 421 AGATTGTTCTGTTAAATATGCTTTATATAAACAAGTTAAAGCTG 467 | |
| Db | 89298 AGATTGTTCTGTTAAATATGCTTTATATAAACAAGTTAAAGCTG 89252 | |
| RESULT 7 | | |
| AF195953/c | | |
| LOCUS | AF195953 | |
| DEFINITION | Homo sapiens membrane-bound aminopeptidase P (XNPEP2) gene, complete cds. | linear PRI 26-MAR-2002 |
| ACCESSION | AF195953 | |
| VERSION | AF195953.2 | GI:19718557 |
| KEYWORDS | | |
| SOURCE | Homo sapiens (human) | |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| AUTHORS | Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C. | |
| TITLE | Human membrane-bound aminopeptidase P genomic DNA | |
| JOURNAL | Unpublished | |
| REFERENCE | 2 (bases 1 to 206618) | |
| AUTHORS | Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C. | |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (18-OCT-1999) Vascular Biology Center, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912, USA | |
| REFERENCE | 3 (bases 1 to 206618) | |
| AUTHORS | Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C. | |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (26-MAR-2002) Vascular Biology Center, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912, USA | |
| REMARK | Sequence update by submitter | |
| COMMENT | On Mar 26, 2002 this sequence version replaced gi:11066156. | |
| FEATURES | Location/Qualifiers | |
| source | 1. .206618 | |
| | /organism="Homo sapiens" | |
| | /mol_type="genomic DNA" | |
| | /db_xref="taxon:9606" | |
| | 144189. .176791 | |
| | /gene="XNPEP2" | |
| | join(144189. .144502,147348. .147421,149203. .149313, | |
| | 150445. .150508,151478. .151582,151837. .151923, | |
| | 152849. .152995,155710. .155811,156987. .157068, | |
| | 157392. .157587,158401. .158490,159714. .159823, | |
| | 160536. .160613,161726. .161797,164422. .164482, | |
| | 165754. .165823,166414. .166518,167248. .167307, | |
| | 167936. .168012,172845. .172934,173533. .176791) | |
| | /gene="XNPEP2" | |
| | /product="membrane-bound aminopeptidase P" | |
| | 144189. .144453 | |
| | /gene="XNPEP2" | |
| | join(144454. .144502,147348. .147421,149203. .149313, | |
| | 150445. .150508,151478. .151582,151837. .151923, | |
| | 152849. .152995,155710. .155811,156987. .157068, | |
| | 157392. .157587,158401. .158490,159714. .159823, | |
| | 160536. .160613,161726. .161797,164422. .164482, | |
| | 165754. .165823,166414. .166518,167248. .167307, | |
| | 167936. .168012,172845. .172934,173533. .176791) | |
| | /gene="XNPEP2" | |
| 5'UTR | | |
| CDS | | |

```
165754. .165823,166414. .166518,167248. .167307,
167936. .168012,172845. .172934,173533. .173727)
/gene="XNPEP2"
/codon_start=1
/product="membrane-bound aminopeptidase P"
/db_xref="GI:11066157"
/translation="MARAHMGCCPWLVLACAWGHTKPVLDGGDVNCSNPPYLP
VTVNTMSLTALQOQNTQNLISAVIIPGTAHMEYIGQDERRAMITGTSAGIA
VVMKKAATWDSYRWTAERQMDCNWELKEVGTPIVWLTTEIPAGGVGDFPFL
LSIDTWESYDALQSNQQLVSIITNLVDLVWGERPPVNPQPIYALQEAFTGWTQE
KVSQVSMQKHQVPTAVLLSALAEETAWLFNLRASDIPYNPFFSYTLTDSIRLF
ANKRFSSETSYLNSCTGPMVCQIEDYSQVRDSIQAYSIGDVRIWIGTSYMYGIV
EMIPEKLIWTVTYSYVMMTKAVKNSKEOALKASHVRDAVAVIRVWLEKNVPKTV
DEFGAETVDFKRFQSPSPSTETISASGLNALAHYSPKELNKLSDENYLLD
SGGQYWDGTTDITRVHGTSPAFQKAYTRVLGNLDSLRIFPAITSGRMVAFRL
RALNDAGLNGHGTHGIGNFLCVHNPVFGFSNNIAWAKMGFTISFPGYIKGEFGL
RLEDVALVVEAKTIPGSLYTFEVVSPVYDRNLIDVLSPEHLQYLNRYQTIREK
VGPBLQRQLLEEFEMLOQHTPELAARAPDTASWASVLVWSTLAILGMSV"
/misc_feature
160606. .160608
/gene="XNPEP2"
/notes="Putative proton shuttle; unclassified site"
/misc_feature
161778. .161780
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
164435. .164437
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
166473. .166475
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
167977. .167979
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
3'UTR
175728. .176791
/gene="XNPEP2"
ORIGIN
Query Match 100.0%; Score 467; DB 9; Length 206618;
Best Local Similarity 100.0%; Pred. No. 2.8e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
DB 51518 CAATGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 51459
QY 61 CTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGGAGCTGGTGGGCGCCAGAA 120
DB 51458 CTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGGAGCTGGTGGGCGCCAGAA 51399
QY 121 GTCTCTTCTGCCACTGACGCCGCCATCAGGAGTTGGGCCTCTTTTCCCTCTCTTCTG 180
DB 51398 GTCTCTTCTGCCACTGACGCCGCCATCAGGAGTTGGGCCTCTTTTCCCTCTCTTCTG 51339
QY 181 TGTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 51338 TGTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 51279
QY 241 GAGAAAGTGGGGATGCTTAAAGAGCTGGGAGATAGGGAACAGAGAGGAGTGGGTG 300
DB 51278 GAGAAAGTGGGGATGCTTAAAGAGCTGGGAGATAGGGAACAGAGAGGAGTGGGTG 51219
QY 301 GGCTAGGGGGCTGCCTATTATTAAGTGTGTTTATGATTCCTTATCTAATTTATACAA 360
DB 51218 GGCTAGGGGGCTGCCTATTATTAAGTGTGTTTATGATTCCTTATCTAATTTATACAA 51159
QY 361 AGATATTAAAGGCTGTTTCAATTAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 420
DB 51158 AGATATTAAAGGCTGTTTCAATTAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 51099
QY 421 AGATGTTCTGCTAAATATATCTTTATATAATAAACAGTTAAAGCTG 467
DB 51098 AGATGTTCTGCTAAATATATCTTTATATAATAAACAGTTAAAGCTG 51052
RESULT 8
AX877684
LOCUS AX877684 2243 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 12589 from Patent EP1074617.
ACCESSION AX877684
VERSION AX877684.1 GI:40032420
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 12589 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2243
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
6..374
/notes="unnamed protein product"
/contig_start=1
/protein_id="CA89834.1"
/db_xref="GI:40032421"
/translation="NSRATWCSPEGQMGQPGREVGNSSAASFPASPIRNPCLSEAGL
XGPPSAHLRLCLLHLVCFSGGLTSLQSPRTCCSHQWALFSPACFPQWRAPGCSL
DDSRSLTRINPVHLPGLSLD"
ORIGIN
Query Match 88.4%; Score 413; DB 6; Length 2243;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGCTGGTGGGTC 114
DB 1831 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGCTGGTGGGTC 1890
QY 115 CAGAAAGTCTCTTCTGCCACTGACGCCGCCATCAGGAGTTGGGCCTTTTCCCTTCC 174
DB 1891 CAGAAAGTCTCTTCTGCCACTGACGCCGCCATCAGGAGTTGGGCCTTTTCCCTTCC 1950
QY 175 TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 234
DB 1951 TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2010
QY 235 GAAGGGGAGAAAGTGGGGGATGGCTTAAAGAGCTGGGAGATAGGGAACAGAGAGGTAG 294
DB 2011 GAAGGGGAGAAAGTGGGGGATGGCTTAAAGAGCTGGGAGATAGGGAACAGAGAGGTAG 2070
QY 295 TGGGTGGGCTAGGGGGCTGCTTATTTAAAGTGTGTTTATGATTCCTTATCTAATTT 354
DB 2071 TGGGTGGGCTAGGGGGCTGCTTATTTAAAGTGTGTTTATGATTCCTTATCTAATTT 2130
QY 355 ATACAAGATATTAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGTTCAATGT 414
DB 2131 ATACAAGATATTAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGTTCAATGT 2190
QY 415 TTGTAAGATGTTCTGCTGCTTAAATATGCTTTTATAATAAACAGTTAAAAAGCTG 467
DB 2191 TTGTAAGATGTTCTGCTGCTTAAATATGCTTTTATAATAAACAGTTAAAAAGCTG 2243
RESULT 9
BD156791
LOCUS BD156791 2243 bp DNA linear PAT 17-JAN-2003
```

DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD156791
VERSION BD156791.1 GI:27862549
KEYWORDS JP 2002191363-A/11634.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2243)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 11634 09-JUL-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/11634
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KACRU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS
Location/Qualifiers
source
1. .2243
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 88.4%; Score 413; DB 6; Length 2243;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 CTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGACTGGTCGGTGC 114
DB 1831 CTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGACTGGTCGGTGC 1890
QY 115 CAGAAAGTCTCTTCTGCACTGACGCCCATCAGGATGGGCTTCTTCCCTTCC 174
DB 1891 CAGAAAGTCTCTTCTGCACTGACGCCCATCAGGATGGGCTTCTTCCCTTCC 1950
QY 175 TTCTGTGCTCTGCTGCTGCTTAAAGTGGTGTGTTATGATCTTTACTAATTT 234
DB 2071 TGGGTGGGCTAGGGGGCTGCTTATTAAGTGGTGTGTTATGATCTTTACTAATTT 2130
QY 355 ATCAAGAATATTAAAGCCCTGTTCAATTAAGAAATGTTCCTTCCCTGTTCAATGT 414
DB 2131 ATCAAGAATATTAAAGCCCTGTTCAATTAAGAAATGTTCCTTCCCTGTTCAATGT 2190
QY 415 TTCTAAAGATTGTTCTGTGTAATATGCTTTTATAATTAACAGTTAAAGCTG 467
DB 2191 TTGTAAGATTGTTCTGTGTAATATGCTTTTATAATTAACAGTTAAAGCTG 2243
RESULT 10
AK001855 2243 bp mRNA linear PRI 01-AUG-2002
LOCUS Homo sapiens cDNA FLJ10993 fis, clone PLACE1002140.
DEFINITION

AK001855
AK001855.1 GI:7023382
oligo capping; fis (full insert sequence).
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2243)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished
REFERENCE 2 (bases 1 to 2243)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
Location/Qualifiers
1. .2243
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1002140"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="cloning vector: pME18SFL3"
6. .374
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91944.1"
/db_xref="GI:7023383"
/translation="MSATWCSPEGGMGQGGPGREVGNSAASGPASPIRNPCLSBAGL KGPSSAHLRLCLLHRLVCFSGSLTSIQLSPRTCSHQWALFSPACFPQWRPGCSL DDSRLSLTRPVLHPGPSLD"
ORIGIN
Query Match 88.4%; Score 413; DB 9; Length 2243;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 CTGCTGCTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGGTCGGTGC 114
DB 1831 CTGCTGCTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGGTCGGTGC 1890
QY 115 CAGAAAGTCTCTTCTGCCACTGACGCCCATCAGGATGGGCTTCTTCCCTTCC 174
DB 1891 CAGAAAGTCTCTTCTGCCACTGACGCCCATCAGGATGGGCTTCTTCCCTTCC 1950
QY 175 TTCTGTGCTCTGCTGCTGCTTAAAGTGGTGTGTTATGATCTTTACTAATTT 234
DB 1951 TTCTGTGCTCTGCTGCTGCTTAAAGTGGTGTGTTATGATCTTTACTAATTT 2010
QY 235 GAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGAAACAGAGAGGTAG 294
DB 2011 GAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGAAACAGAGAGGTAG 2070
QY 295 TGGGTGGGCTAGGGGGCTGCTTATTAAGTGGTGTGTTATGATCTTTACTAATTT 354
DB 2071 TGGGTGGGCTAGGGGGCTGCTTATTAAGTGGTGTGTTATGATCTTTACTAATTT 2130
QY 355 ATCAAGAATATTAAAGCCCTGTTCAATTAAGAAATGTTCCTTCCCTGTTCAATGT 414
|||||

```

Db      2131 ATACAAAGATATTAAGGCCCTGTTTCATTAAGAAATGTTCCCTCCCTGTTGTTCAATGT 2190
QY      415 TTGTAAGATGTTCTGTGTAATAATGCTTTATATAACAGTTAAAGCTG 467
Db      2191 TTGTAAGATGTTCTGTGTAATAATGCTTTATATAACAGTTAAAGCTG 2243

RESULT 11
LOCUS   AX332625/c
DEFINITION Sequence 3134 from Patent WO0194629.
ACCESSION AX332625
VERSION   AX332625.1 GI:18123259
KEYWORDS
SOURCE  Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
          Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE    Cancer gene determination and therapeutic screening using signature
          gene sets
JOURNAL  Avalon Pharmaceuticals (US)
FEATURES
source   Location/Qualifiers
          1..458
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.6%; Score 241; DB 6; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.4e-127;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 286
Db 252 CCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 193
QY 287 GAGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 346
Db 192 GAGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 133
QY 347 ACTAATTTATACAAAGATATTAAGGCCCTGTTTATTAAGAAATGTTCCCTCCCTGTG 406
Db 132 ACTAATTTATACAAAGATATTAAGGCCCTGTTTATTAAGAAATGTTCCCTCCCTGTG 73
QY 407 TTCAATGTTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATATAACAGTTAAAGCT 466
Db 72 TTCAATGTTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATATAACAGTTAAAGCT 13
QY 467 G 467
Db 12 G 12

RESULT 12
LOCUS   AX332852/c
DEFINITION Sequence 3361 from Patent WO0194629.
ACCESSION AX332852
VERSION   AX332852.1 GI:18123486
KEYWORDS
SOURCE  Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
          Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE    Cancer gene determination and therapeutic screening using signature
          gene sets
JOURNAL  Avalon Pharmaceuticals (US)
FEATURES
source   Location/Qualifiers
          1..458
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.6%; Score 241; DB 6; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.4e-127;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 286
Db 252 CCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 193
QY 287 GAGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 346
Db 192 GAGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 133
QY 347 ACTAATTTATACAAAGATATTAAGGCCCTGTTTATTAAGAAATGTTCCCTCCCTGTG 406
Db 132 ACTAATTTATACAAAGATATTAAGGCCCTGTTTATTAAGAAATGTTCCCTCCCTGTG 73
QY 407 TTCAATGTTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATATAACAGTTAAAGCT 466
Db 72 TTCAATGTTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATATAACAGTTAAAGCT 13
QY 467 G 467
Db 12 G 12

RESULT 13
LOCUS   AX874329/c
DEFINITION Sequence 9234 from Patent EP1074617.
ACCESSION AX874329
VERSION   AX874329.1 GI:40029100
KEYWORDS
SOURCE  Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ota, T., Isoqai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
          Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE    Primers for synthesising full-length cDNA and their use
JOURNAL  Patent: EP 1074617-A 9234 07-FEB-2001;
          Research Association for Biotechnology (JP)
FEATURES
source   Location/Qualifiers
          1..403
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.0%; Score 238; DB 6; Length 403;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 TGTGTCTCTCTCATCGGCTGCCTGACCTGACCTGAGCAAGCCAGCCCGCTGGGGAAG 238
Db 289 TGTGTCTCTCTCATCGGCTGCCTGACCTGAGCAAGCCAGCCCGCTGGGGAAG 230
QY 239 GGGAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGG 298
Db 229 GGGAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGG 170
QY 299 TGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTATATAATATAC 358
Db 169 TGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTATATAATATAC 110

```

JOURNAL Patent: WO 0194629-A 3361 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
source Location/Qualifiers
1..458
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 51.6%; Score 241; DB 6; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.4e-127;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 227 CCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 286
Db 252 CCCGTGGGGAAGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 193
QY 287 GAGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 346
Db 192 GAGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 133
QY 347 ACTAATTTATACAAAGATATTAAGGCCCTGTTTATTAAGAAATGTTCCCTCCCTGTG 406
Db 132 ACTAATTTATACAAAGATATTAAGGCCCTGTTTATTAAGAAATGTTCCCTCCCTGTG 73
QY 407 TTCAATGTTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATATAACAGTTAAAGCT 466
Db 72 TTCAATGTTTGTAAAGATGTTCTGTGTAATAATGTTCTTTATATAACAGTTAAAGCT 13
QY 467 G 467
Db 12 G 12

```

RESULT 13

AX874329/c
LOCUS AX874329 403 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 9234 from Patent EP1074617.
ACCESSION AX874329
VERSION AX874329.1 GI:40029100
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ota, T., Isoqai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 9234 07-FEB-2001;
Research Association for Biotechnology (JP)

FEATURES
source Location/Qualifiers
1..403
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 51.0%; Score 238; DB 6; Length 403;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 179 TGTGTCTCTCTCATCGGCTGCCTGACCTGACCTGAGCAAGCCAGCCCGCTGGGGAAG 238
Db 289 TGTGTCTCTCTCATCGGCTGCCTGACCTGAGCAAGCCAGCCCGCTGGGGAAG 230
QY 239 GGGAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGG 298
Db 229 GGGAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGG 170
QY 299 TGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTATATAATATAC 358
Db 169 TGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTATATAATATAC 110

```


Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23913
 Center clone name: 406_O_24
 ----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 199596 bases at least Q40
 Consensus quality: 200632 bases at least Q30
 Consensus quality: 201122 bases at least Q20
 Insert size: 183000; agarose-fp
 Quality coverage: 10.7 in Q20 bases; sum-of-contigs
 Quality coverage: 9.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 8997: contig of 8997 bp in length
 * 8998 9097: gap of 100 bp
 * 9098 9921: contig of 824 bp in length
 * 9922 10021: gap of 100 bp
 * 10022 11236: contig of 1215 bp in length
 * 11237 11336: gap of 100 bp
 * 11337 12662: contig of 1326 bp in length
 * 12663 12762: gap of 100 bp
 * 12763 15814: contig of 3052 bp in length
 * 15815 15914: gap of 100 bp
 * 15915 27843: contig of 11929 bp in length
 * 27844 27943: gap of 100 bp
 * 27944 100820: contig of 72877 bp in length
 * 100821 172522: gap of 100 bp
 * 100921 172522: contig of 71602 bp in length
 * 172523 172622: gap of 100 bp
 * 172623 202198: contig of 29576 bp in length.

FEATURES

source

1. 202198
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="np23-406024"
 /clone_lib="RPCI-23 Female Mouse BAC"
 misc_feature
 1. 8997
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 misc_feature
 9098..9921
 /note="assembly_fragment"
 misc_feature
 10022..11236
 /note="assembly_fragment"
 misc_feature
 11337..12662
 /note="assembly_fragment"
 misc_feature
 12763..15814
 /note="assembly_fragment"
 misc_feature
 15915..27843
 /note="assembly_fragment"
 misc_feature
 27944..100820
 /note="assembly_fragment"
 misc_feature
 100921..172522
 /note="assembly_fragment"
 misc_feature
 172623..202198
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right

RESULT 16

AC122549

LOCUS

DEFINITION

AC122549

AC122549

HTG: HTGS PHASE1; HTGS_DRAFT.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 202198)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-406024

Unpublished

2 (bases 1 to 202198)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,

Boukhgalter,B., Brown,A., Camarata,J., Campionano,A., Chang,J.,

Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,

Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., Fitzgerald,M., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,

Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lakocque,K.,

Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K.,

Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,

Mathews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,

Meneus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C.,

Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Oliver,J., Peterson,K., Phukhang,P., Pierre,N.,

Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,

Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,

Schupbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,

Theodore,J., Topham,K., Travers,M., Travis,N., Triggilio,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zembek,V., Zimmer,A. and Zody,M.

Young,G., Zainoun,J., Zembek,V., Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 202198)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,

Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,

Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K.,

Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,

Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,

Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,

Meldrum,J., Meneus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J.,

Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,

O'Neil,D., Oliver,J., Peterson,K., Phukhang,P., Pierre,N.,

Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,

Roman,J., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Smith,C.,

Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,

Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,

Vassiliev,H., Venkataran,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,

Wyman,D., Young,G., Zainoun,J., Zembek,V., Zimmer,A. and Zody,M.

Wyman,D., Young,G., Zainoun,J., Zembek,V., Zimmer,A. and Zody,M.

Direct Submission

Submitted (13-MAR-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 13, 2003 this sequence version replaced gi:28261613.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

ORIGIN

Query Match 6.0%; Score 28; DB 2; Length 202198;
 Best Local Similarity 100.0%; Pred. No. 0.00031;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 50 TAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 77
 Db 104108 TAAAGCTGCTGCTGCTGCTGCTGCTGCT 104135

RESULT 17

AC131317 204171 bp DNA linear HTG 21-MAR-2003
 LOCUS Mus musculus clone RP23-333F2, WORKING DRAFT SEQUENCE, 9 unordered
 DEFINITION pieces.
 AC131317
 AC131317.4 GI:29135648
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Mus musculus (house mouse)
 SOURCE
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 204171)
 Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-333F2

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

COMMENT

On Mar 21, 2003 this sequence version replaced gi:28412031.

All repeats were identified using RepeatMasker:

Snit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: 126954

Center clone name: 333_F_2

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 202347 bases at least Q40

Consensus quality: 202846 bases at least Q30

Consensus quality: 202986 bases at least Q20

Insert size: 198000; agarose-fp

Quality coverage: 8.9 in Q20 bases; agarose-fp

Quality coverage: 8.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 61474: contig of 61474 bp in length

* 61475 61574: gap of 100 bp

* 61575 64479: contig of 2905 bp in length

* 64480 64579: gap of 100 bp

* 64580 69264: contig of 4685 bp in length

* 69265 69364: gap of 100 bp

* 69365 85451: contig of 16087 bp in length

* 85452 105030: contig of 19479 bp in length

* 105031 105130: gap of 100 bp

* 105131 124412: contig of 19282 bp in length

* 124413 124512: gap of 100 bp

* 124513 143983: contig of 19471 bp in length

* 143984 144083: gap of 100 bp

* 144084 202781: contig of 58698 bp in length

* 202782 202881: gap of 100 bp

* 202882 204171: contig of 1290 bp in length.

* Location/Qualifiers

1..204171

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/clone="RP23-333F2"

/clone_lib="RPCI-23 Female Mouse BAC"

1..61474

/note="assembly_fragment"

clone_end:sp6

vector_side:left

61575..64479

/note="assembly_fragment"

64580..69264

/note="assembly_fragment"

69365..85451

/note="assembly_fragment"

85452..105030

/note="assembly_fragment"

105131..124412

/note="assembly_fragment"

124513..143983

/note="assembly_fragment"

144084..202781

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature
    /note="assembly_fragment"
    202882..204171-
    /note="assembly_fragment"
    clone_end:17
    vector_side:right"

ORIGIN
Query Match          6.0%; Score 28; DB 2; Length 204171;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAAA 81
      |||
Db 53830 GCTGCTGCTGCTGCTGCTGCTTAAA 53857

RESULT 18
AY023222
LOCUS AY023222 227 bp DNA linear PLN 07-FEB-2001
DEFINITION Oryza sativa microsatellite MRS5547 containing (GCT)X9, closest to
marker L825, genomic sequence.
ACCESSION AY023222
VERSION AY023222.1 GI:12706438
KEYWORDS
SOURCE
ORGANISM Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 227)
Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
Simple sequence repeats from Monsanto rice genomic sequences
Unpublished
REFERENCE 2 (bases 1 to 227)
Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
Direct Submission
Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
Blvd., Creve Coeur, MO 63167, USA
Derived from rice genomic sequences generated from the Monsanto
Rice Genome Sequencing project. Please see
http://www.rice-research.org for more information. The sequence
data were produced primarily in the laboratories of Dr. Leroy Hood
at the University of Washington in Seattle.
FEATURES
source
1..227
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
repeat_region
1..227
/note="microsatellite MRS5547"
/rpt_type=tandem
/rpt_unit="gct"

ORIGIN
Query Match          5.8%; Score 27; DB 8; Length 227;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
      |||
Db 98 AACGCTGCTGCTGCTGCTGCTGCTGCT 124

RESULT 19
G09506
LOCUS G09506 373 bp DNA linear STS 14-AUG-1995
DEFINITION human STS CHLC.GCT4G01.P10966 clone GCT4G01, sequence tagged site.
ACCESSION G09506
VERSION G09506.1 GI:941355
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
TITLE Cooperative Human Linkage Center
JOURNAL Unpublished (1995)
COMMENT Synonyms: GCT4G01, CHLC.GCT4G01.T10965
Contact: Dr. Jeffrey C. Murray
UoI The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jeff-murray@uiowa.edu

Primer A: TCGGGTAAAGATGAATG
Primer B: ACTCTAGGAATGAGACATTCG
STS size: 106
PCR Profile:
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degrees C

Protocol:
Template: 30ng genomic DNA
Primer: each 1.5 pmole
dNTPs: each 200 uM
Taq Polymerase: 0.3 units
Total Vol: 10 ul

Buffer:
MgCl2: 1.5mM
KCl: 50mM
Tris: 10mM
pH: 8.3.
Location/Qualifiers
1..373
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
12..117
primer_bind
12..31
primer_bind
complement(96..117)

Query Match          5.8%; Score 27; DB 11; Length 373;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCTTA 79
      |||
Db 69 CGCTGCTGCTGCTGCTGCTGCTGCTTA 95

RESULT 20
AB028183/c
LOCUS AB028183 1264 bp mRNA linear PLN 22-JAN-2000
DEFINITION Oryza sativa mRNA for OsNAC4 protein, complete cds.
ACCESSION AB028183
VERSION AB028183.1 GI:6730937
KEYWORDS OsNAC4; OsNAC4 protein.
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (sites)
Kikuchi,K., Ueguchi-Tanaka,M., Yoshida,K.T., Nagato,Y.,
Matsusaka,M. and Hirano,H.Y.
Molecular analysis of the NAC gene family in rice
Mol. Gen. Genet. 262 (6), 1047-1051 (2000)
MEDLINE 20123461
PUBMED 10660065

```


The orientation of the sequence is from -21M13 to M13rev of the BAC clone.

```

FEATURES             source
    source
    1. 88088
    /organism="Oryza sativa (japonica cultivar-group)"
    /mol_type="genomic DNA"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosome="1"
    /clone="OSUNBa0085D07"

ORIGIN
Query Match          5.8%; Score 27; DB 8; Length 88088;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 55199 AACGCTGCTGCTGCTGCTGCTGCTGCT 55225

RESULT 25
AC017472
LOCUS               AC017472 128403 bp DNA linear HTG 09-DEC-1999
DEFINITION          Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION            AC017472
VERSION              AC017472.1 GI:6553514
KEYWORDS             HTG; HTGS PHASE2.
SOURCE               Drosophila melanogaster (fruit fly)
ORGANISM             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                    Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                    Ephydroidea; Drosophilidae; Drosophila.
REFERENCE            Adams,M. and Venter,J.C.
AUTHORS              Submitted
TITLE               Direct Submission
JOURNAL              Rockville, MD, USA
COMMENT              This sequence was identified as CDM:10210908 by the submitter.
                    For more information on this record e-mail to fly@celera.com.
                    * NOTE: This is a 'working draft' sequence.
                    * This sequence will be replaced
                    * by the finished sequence as soon as it is available and
                    * the accession number will be preserved.
FEATURES             source
    Location/Qualifiers
    1. 128403
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"

ORIGIN
Query Match          5.8%; Score 27; DB 2; Length 128403;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCTGCT 79
Db 91341 CGCTGCTGCTGCTGCTGCTGCTGCTGCT 91367

RESULT 26
AP003411
LOCUS               AP003411 132059 bp DNA linear PLN 27-NOV-2003
DEFINITION          Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
                    BAC clone: B1148D12.
ACCESSION            AP003411 BA000010.
VERSION              AP003411.3 GI:15528778
KEYWORDS             Oryza sativa (japonica cultivar-group)
ORGANISM             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                    Ehrhartoideae; Oryzeae; Oryza.

```

REFERENCE

AUTHORS

1 Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T., Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y., Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K., Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K., Hanada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y., Idonuma,A., Iijima,W., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T., Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S., Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T., Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M., Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M., Ohra,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T., Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsujii,K., Waki,K., Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,K., Yukawa,K., Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y., Yano,M., Jiang,J. and Gojobori,T.
The genome sequence and structure of rice chromosome 1 Nature 420 (6913), 312-316 (2002)
22337376
12447438
2 (bases 1 to 132059)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (14-MAR-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kamondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
On Sep 7, 2001 this sequence version replaced gi:14624988.
Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI NonRedundant protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.
The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of B1148D12 clone has an overlap with P0676H03 (DDBJ: AP003285) clone at the position 1 to 45,178 and with P0454H12 (DDBJ: AP003255) at the position 56,068 to 132,059 of 5' end. The sequence of this clone starts at the position 117,675 of P0676H03 and ends at the position 75,992 of P0454H12. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
Location/Qualifiers
1. 132059
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="B1148D12"
complement(join(1616..2083,2184..2470,2569..2764))
/gene="B1148D12.1"
complement(join(1616..2083,2184..2470,2569..2764))
/gene="B1148D12.1"
/note="Contains ESTs AU075972 (C10012), D21888 (C10012)"
/codon_start=1
/product="development regulation gene OsNAC4"
/protein_id="BAB64820.1"
/db_xref="GI:15528779"
/translation="MAAVGGSGRRDAELNLPGRFPHPTDEEIVHYLCRKVARQ PLPVPITAEVDLYKLDPEKALFGKWEYFFTPDRKYPNGSRPNRAGRGYKA

FEATURES

source

gene

CDS


```

TGADKPVAKGSAITGVIKALVYSGKAPGVKTDWIMTHEYRLADADRACGCKGSGQ
KLDWVLCRLYNKKNNEKYKLEQDDVAVSAAAPRNHHQNGVMDRAADINSDSF
QTHSDIDNAGAHGCGGGGFGDVPAPPPNGVTVKEDNDWTFGLNFDELOPPYMM
NLHQMOMNAPAPAGDDGGYQISISPPGMMQWOTLPPFF"
join(6090) .6317,6421. .6572,8325. .8440,8715. .8803,
9502. .9589,9669. .9774,10773. .10866,11020. .11192,
11536. .11575,12734. .12875,13216. .13303,13676. .13730)
/gene="B1148D12.2"
/gene="B1148D12.3"
join(6090) .6317,6421. .6572,8325. .8440,8715. .8803,
9502. .9589,9669. .9774,10773. .10866,11020. .11192,
11536. .11575,12734. .12875,13216. .13303,13676. .13730)
/gene="B1148D12.2"
/note="contains EST"
hypothetical protein"
/codon_start=1
/protein_id="BAB64821.1"
/db_xref="GI:15528780"
/translating="MRAACEGCGMWTAAAGLASRWGTSAGSGGSLRELNHANAVA
TSFLRHRSQLRWEGQDAKRGTVGNVAFSEASMSPEYAPSPSIALPDPPEASRV
ACRLYHRSQCPHGLFLLPARGHACTHALTGNPQPAVGTGYTLACGNCPCDYC
RVASQNTVRMSDNCASRTVVGSTPRQEDTQGTGSLSPVTFDRIGGCAPNFHPV
PNGVAGGVGLMSVSGGRWGEDEGRLLSFQAKAQIVVRASCGYWSRTPPLAAT
ESQDDVTRPRRGEHVEHLCYXYGVRNVTVLSTMHPRPCCLNSRTSSMIGS
SMSGFRGFRFSRQIKKVDYSVTGDQLHVELAHWENFHRHRTAAAGGSQFSYVVA
ESQCRPNTAAGLHRTVFSRLVAYYTCNVCYLRCSLYVDKCSRTPLPLRKQVH
LHLSGK"
complement(join(14359) .14600,14752. .14929,15013. .15084,
15363. .15458,15562. .15678,15760. .15916,16010. .16107,
16180. .16338,16409. .16567,16788. .16907,17208. .17325,
17456. .17586,17692. .17768,18098. .18209,18297. .18479,
18656. .18757,18857. .18958,19057. .19152,19721. .19798,
19892. .19984,20057. .20178,20275. .20356,20444. .20518,
20625. .20885,20970. .21094,21196. .21271,21368. .21517,
21599. .21678,21764. .21865,21947. .22054,23134. .23246,
23518. .23623,24383. .24487,24596. .24667,24690. .24847,
25307. .25415,25492. .25557,25644. .25780,26028. .26130,
26656. .26775,26947. .27060,27163. .27248,27328. .27410,
28523. .28587,28684. .28861,28937. .29016,30166. .30289,
30380. .30485,30639. .30801)
/gene="B1148D12.3"
complement(join(14359) .14600,14752. .14929,15013. .15084,
15363. .15458,15562. .15678,15760. .15916,16010. .16107,
16180. .16338,16409. .16567,16788. .16907,17208. .17325,
17456. .17586,17692. .17768,18098. .18209,18297. .18479,
18656. .18757,18857. .18958,19057. .19152,19721. .19798,
19892. .19984,20057. .20178,20275. .20356,20444. .20518,
20625. .20885,20970. .21094,21196. .21271,21368. .21517,
21599. .21678,21764. .21865,21947. .22054,23134. .23246,
23518. .23623,24383. .24487,24596. .24667,24690. .24847,
25307. .25415,25492. .25557,25644. .25780,26028. .26130,
26656. .26775,26947. .27060,27163. .27248,27328. .27410,
28523. .28587,28684. .28861,28937. .29016,30166. .30289,
30380. .30485,30639. .30801)
/note="contains EST C25950(c11258)"
/codon_start=1
/product="putative microtubule associated protein"
/protein_id="BAB64822.1"
/db_xref="GI:15528781"
DPKDALREFGPIFKTVDASNPVQEKALDALLAFORAADADASYAKEVCDAIVAK
CLTGRPTVKAQAQAFLLWYDEAAEVLSEKAVKAKVAVPAIDVFMQALSEF
GAKVVPKKLKMILPELDHPQDNVRASSKGLTLELCRWIGKEPVKAILPEKRWDMK
KELEAVLNSGIAKPRKIRSEOELEEVPEAAGTNNSEAVPEAPMEIDEVYL
VDVDDILTPLEKSGFVGKATKWSERDAVELTKLSTKAPGDFHEICRTLKL
ITDNLAVSVEAQIAGNLKRLHSGNSRVLPLVLEKKEKKEPTTEAUSQTLQ
AMHSGCITLLEDVVRVAKNKPVLRSITLNNWAFCTISNKAITVLKHXEYPI
CMELNDGTEPVSFSLVTAIKVMGKPLERLEKLDVVRKKLSMDIGASDIT
SGTESFQIDIIPEFGKEMCRATVVPVTLNDYELLARGRLVAFVLOTADPTVM
DSSSMRASAASMLSGKPPQVAPATKSPAKSATAKTKDGGPQSKASAPVIEDVEP
SEMSLEIEEKLSSVVKSETISQKSTVWKERLEASMLKQVESITELDKSRELLVR
LLCAVPQWSENQVLIKIMKHKNPKVLSIGLLMWSAVEDFGISLNUKOTIDFCO
IGLQSSAAATRNATIKLIGVILHFKVGVGDIKGLFSLDKVPALLSTLDAIEYKNNPEGTAS

```

```

APRTRVRAADAVSSASGTSDDGLPREDISAKITPTLLKLNIGSPWKLRLSIEDAVSKI
VEBAHQIQTGTVELFTALRLYDSNKNLVNATLSTIGGLASGAPFAVEKSGKIL
ADVLKCLGDNKEMRECTIATLIDLVAAQAOLDKMVPYITVTLGQTKTQKXOLFOW
LSKHANMSPSEALPLPKPSASSLMDKSSKSEVRAKAAESFNNEILKICQDQVAKNLO
LPSTLAIVAERLKLSTVHGFSDSVKMTTSMLSKPSKAGSKNNKHGPNDRGNSVKA
VSQGIPTARSVTMISDQSIQOALFNIKDSNKEERERVLVRKFEFEPREPRQIDE
LKLELFHFRFEDVSLRLWNSDFRQIDQIELLOKALPSSRERVELLIDILLRWFVLR
CESNTCLLKLVDLPFLFDVKDQSYMLTEAAALFLPCMSEKSGNEVSGLLKPMGE
LIFQMNITYSLKPLLYLEGLRSKNNRTIECVDIITGYFMDHGHGTSVGLLKPMGE
AALTARDEGEIRKALNTLATAYKNLGDVVRVVGKLSDAQRSMLDDRFKWKAREMDK
RRGPPCARAALRRSVRENGSDIAEGSGEAVRSAGENFGYSDAMVPRMATATP
GDMREALDVALGPEQSGEKMVICHELTONVDPRESSVLDLKEADRLVSCILAV
MTQIKRLAHAVKEGILNPSITPPESLAVKNOKFSDLVVKCLIKLTKVQSTIYEV
RTSFFVLNLLRPDPSRWPSPTPPESLAVKNOKFSDLVVKCLIKLTKVQSTIYEV
DLDRILQSTHIYLOELGMEIRRRAGADKPLRMVKVTLHELKLGRTAIKGLHSMVP
IDAEPOPITILAYDLNLQTLAAARMLTPSGTGTGTGWDAGSNPNPSTHSTDLQKQ
ELAAFKKIGDKOTCTIGLYELVITQLYPKVDIFALQNASAFRTYIRDLAQAQVK
NAAAGTPSSLELSTPPIAPISPKFAPSVPHTKSINNKTCDCNEDDAGDGTPEFCT
LDALRRMSKIQAAAVGANFDGVQAPRPLSMNGNTLHGTRLDADPTQNIIFPMDBR
ALSGLQARMERLXSGSMEPL"
complement(41340) .43701)
/gene="B1148D12.4"
complement(41340) .43701)
/note="contains EST"
probably inactive due to no termination codon in CDS
pseudogene
similar to Arabidopsis thaliana chromosome 2, At2g35630"
/pseudo
join(46995) .47010,47163. .47331,48588. .48659,48754. .48915,
49020. .49094,49172. .49243,49329. .49400,49485. .49553,
49972. .50106,50186. .50655,50730. .50888,50981. .51120,
51214. .51433,51545. .51780,51867. .52370)
/gene="B1148D12.5"
join(46995) .47010,47163. .47331,48588. .48659,48754. .48915,
49020. .49094,49172. .49243,49329. .49400,49485. .49553,
49972. .50106,50186. .50655,50730. .50888,50981. .51120,
51214. .51433,51545. .51780,51867. .52370)
/note="contains EST"
/codon_start=1
/product="putative receptor protein kinase"
/protein_id="BAB64823.1"
Query Match 5.8%; Score 27; DB 8; Length 132059;
Best Local Similarity 100.0%; Pred.No.0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 2817 AACGCTGCTGCTGCTGCTGCTGCTGCT 2843
|||||
RESULT 27
AC103357 143558 bp DNA linear HTG 04-DEC-2003
LOCUS Mus musculus chromosome 14 clone RP24-201114 map 14, *** SEQUENCING
IN PROGRESS ***, 2 ordered pieces.
AC103357
AC103357.7 GI:38678639
HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS Mus musculus
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 143558)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus chromosome 14, clone RP24-201114
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143558)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,I., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

```



```
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-1316"
/clone_lib="Medicago truncatula BAC library H2"
```

ORIGIN

```
Query Match      5.8%; Score 27; DB 2; Length 158606;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 50 TAACGCTGCTGCTGCTGCTGCTGCTGCTG 76
|||||
Db 84940 TAACGCTGCTGCTGCTGCTGCTGCTGCTG 84914
```

RESULT 29

```
AC010921/c
LOCUS      163466 bp DNA linear INV 14-MAR-2001
DEFINITION Drosophila melanogaster, chromosome X, region 15B-15E, BAC clone
BACR15L12, complete sequence.
```

```
AC010921
```

```
AC010921.11 GI:13324755
```

```
HTG.
```

```
SOURCE    Drosophila melanogaster (fruit fly)
```

```
ORGANISM  Drosophila melanogaster
```

```
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
```

```
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
Ephydroidea; Drosophilidae; Drosophila.
```

```
1 (bases 1 to 163466)
```

```
REFERENCE
AUTHORS    Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busan,D.A.,
Carlson,J.W., Center,A., Chape,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattai,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome X, region 15B-15E
Unpublished
```

```
2 (bases 1 to 163466)
```

```
REFERENCE
AUTHORS    Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champetier,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
```

```
Direct Submission
```

```
Submitted (27-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
```

```
On Mar 14, 2001 this sequence version replaced gi:7143399.
```

```
Sequence submitted by:
```

```
Berkeley Drosophila Genome Project
```

```
Lawrence Berkeley National Laboratory, MS 64-121
```

```
Berkeley, CA 94720
```

```
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
```

FEATURES

```
source
```

FEATURES

```
source
```

```
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="15B-15E"
/clone="BACR15L12 (D898)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
```

ORIGIN

```
Query Match      5.8%; Score 27; DB 3; Length 163466;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 53 CGCTGCTGCTGCTGCTGCTGCTGCTGCTTA 79
|||||
Db 94716 CGCTGCTGCTGCTGCTGCTGCTGCTTA 94690
```

RESULT 30

```
AL954848
```

```
LOCUS
```

```
DEFINITION Zebrafish DNA sequence from clone CH211-281G7, complete sequence.
```

```
AL954848
```

```
ACCESSION
```

```
VERSION
```

```
AL954848.9 GI:29165221
```

```
HTG.
```

```
KEYWORDS
```

```
SOURCE
```

```
ORGANISM
```

```
Danio rerio (zebrafish)
```

```
Danio rerio
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
```

```
Cypriniformes; Cyprinidae; Danio.
```

```
1 (bases 1 to 170178)
```

```
Johnson,C.
```

```
Direct Submission
```

```
Submitted (23-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
```

```
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
```

```
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
On Mar 23, 2003 this sequence version replaced gi:29125265.
```

```
----- Genome Center
```

```
Center: Wellcome Trust Sanger Institute
```

```
Center code: SC
```

```
Web site: http://www.sanger.ac.uk
```

```
Contact: zfish-help@sanger.ac.uk
```

```
-----
```

```
During sequence assembly data is compared from overlapping clones.
```

```
Where differences are found these are annotated as variations
```

```
together with a note of the overlapping clone name. Note that the
```

```
variation annotation may not be found in the sequence submission
```

```
corresponding to the overlapping clone, as we submit sequences with
```

```
only a small overlap as described above.
```

```
This sequence was finished as follows unless otherwise noted: all
```

```
regions were either double-stranded or sequenced with an alternate
```

```
chemistry or covered by high quality data (i.e., phred quality >=
```

```
30); an attempt was made to resolve all sequencing problems, such
```

```
as compressions and repeats; all regions were covered by at least
```

```
one plasmid subclone or more than one M13 subclone; and the
```

```
assembly was confirmed by restriction digest, except on the rare
```

```
occasion of the clone being a YAC.
```

```
The following abbreviations are used to associate primary accession
```

```
numbers given in the feature table with their source databases:
```

```
Em: ENBL; Sw: SWISSPROT; Tr: TRMBL; Wp: WORMPEP; Information
```

```
on the WORMPEP database can be found at
```

```
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep Repeat names
```

```
beginning 'Dr' were identified by the Recon repeat discovery system
```

```
(Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
```

```
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
```

```
further information see http://www.projects/d\_rerio/fishmask.shtml
```

```
CH211-281G7 is from a CHORI-211 BAC library
```

```
VECTOR: pTABAC2.1
```

```
Location/Qualifiers
```

```
1..170178
```

```
/organism="Danio rerio"
```

```
/mol_type="genomic DNA"
```

```

/db_xref="taxon:7955"
/clonelib="CH211-281G7"
/clonelib="CHORI-211"

ORIGIN
Query Match      5.8%; Score 27; DB 5; Length 170178;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AGCTGCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 36000 AGCTGCTGCTGCTGCTGCTGCTGCTT 36026

RESULT 31
AC012160 172069 bp DNA linear INV 13-DEC-2001
Drosophila melanogaster, chromosome X, region 15E-15E, BAC clone
BAC06G02, complete sequence.
AC012160
AC012160.6 GI:17646875
HTG
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 172069)
REFERENCE
AUTHORS
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferrera,S., Frise,B., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Hocky,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Izbagwan,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome X, region 15E-15E
Unpublished (1998)
2 (bases 1 to 172069)
Celniker,S.E., Agbavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snitz,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 13, 2001 this sequence version replaced gi:6466928.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
1. .172069
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"

FEATURES
Source

```

```

/chromosome="X"
/map="15E-15E"
/clonelib="BACR06G02 (D1110)"
/clonelib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACE3.6)"

ORIGIN
Query Match      5.8%; Score 27; DB 3; Length 172069;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGTGTGCTGCTGCTGCTGCTGCTGCTTA 79
Db 8990 CGTGTGCTGCTGCTGCTGCTGCTGCTTA 9016

RESULT 32
AC137828 172332 bp DNA linear HTG 10-DEC-2003
Medicago truncatula clone mth2-6c16, WORKING DRAFT SEQUENCE, 15
unordered pieces.
AC137828
AC137828.15 GI:39652644
HTG; HTGS PHASE1; HTGS DRAFT.
Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 172332)
Shaul,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Medicago truncatula BAC Clone mth2-6c16
Unpublished
2 (bases 1 to 172332)
Shaul,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (03-DEC-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 172332)
Shaul,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (10-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Dec 10, 2003 this sequence version replaced gi:39228232.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2136: contig of 2136 bp in length
* 2137 2236: gap of unknown length
* 2237 4244: contig of 2008 bp in length
* 4245 4345: gap of unknown length
* 4345 7875: contig of 3531 bp in length
* 7876 10848: gap of unknown length
* 7976 10848: contig of 2873 bp in length
* 10849 14384: gap of unknown length
* 10949 14384: contig of 3436 bp in length

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

6 (bases 1 to 177062)
Wilson,R.
Direct Submission
Submitted (11-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 29, 2002 this sequence version replaced gi:22476077.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0198P16
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC112268.
Location/Qualifiers
1..177062
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/map="2"
/clone="RP24-198P16"
/clone_lib="RPCI-24"
8..42
/repeat_region
/repeat_family="Alu"
220..619
/repeat_region
/repeat_family="MaLR"
767..992
/repeat_region
/rpt_family="ERVK"
1228..1462
/rpt_family="ERVL"
2151..2375
/rpt_family="ERV1"
2692..3019
/rpt_family="MaLR"
4693..4845
/rpt_family="MaLR"
4846..4975
/rpt_family="L1"
5183..5301
/rpt_family="Alu"
5382..5453
/rpt_family="Alu"
5494..7230
/repeat_region

```

```

/rpt family="L1"
7388..7448
/rpt family="L1"
7684..7851
/rpt family="B2"
7853..8183
/rpt family="L1"
8189..8438
/rpt family="L1"
8439..8646
/rpt family="B2"
8647..9273
/rpt family="L1"
9298..9577
/rpt family="B4"
9580..10116
/rpt family="L1"
10961..11004
/rpt family="L1"
11022..11154
/rpt family="Alu"
11283..11430
/rpt family="L1"
13190..13397
/rpt family="B2"
13575..13721
/rpt family="Alu"
14531..14824
/rpt family="RMER10A"
15075..15141
/rpt family="L1"
15179..15313
/rpt family="B4"
15433..15695
/rpt family="L1"
20710..20873
/Note="Sequence derived from one plasmid subclone."
repeat_region
22130..22261
/rpt family="B4"
22592..22806
/rpt family="L1"
23136..23225
/rpt family="Alu"
23456..23516
/rpt family="ID"
24444..24504
/rpt family="Alu"
24653..24822
/rpt family="B2"
25552..25755
/rpt family="B2"
27915..27973
/rpt family="ERV1"
28250..28283
/rpt family="B2"
29135..29329
/rpt family="B2"
31052..31177
/rpt family="Alu"
31695..31841
/rpt family="MER1_type"
32462..32583
/rpt family="B4"
32785..32839
/rpt family="ID"
34095..34169
/rpt family="Tip100"
35610..35998
/rpt family="MaLR"
36012..36274
/rpt family="MaLR"
37866..37895
/rpt family="MaLR"

```

```

repeat_region 38110..38286
/rpt family="B2"
repeat_region 38315..38427
/rpt family="B2"
repeat_region 38707..38775
/rpt family="ERV1"
repeat_region 41034..41175
/rpt family="Alu"
repeat_region 42647..43068
/rpt family="L1"
repeat_region 43072..44040
/rpt family="L1"
repeat_region 44182..44313
/rpt family="B2"
repeat_region 44381..44438
/rpt family="ERV1"
repeat_region 44495..44556
/rpt family="L1"
repeat_region 44557..44744
/rpt family="B2"
complement(44666..44738)

Query Match 5.8% Score 27; DB 10; Length 177062;
Best Local Similarity 100.0%; Pred.No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 110056 AACCTGCTGCTGCTGCTGCTGCT 110082

RESULT 34
AC102553/c
LOCUS AC102553 185722 bp DNA linear HTG 18-JUN-2003
DEFINITION Mus musculus clone RP23-159D6, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
AC102553
AC102553.3 GI:31880154
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 185722)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-159D6
Unpublished
REFERENCE 2 (bases 1 to 185722)
Birren,B., Linton,J., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,M., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

REFERENCE
AUTHORS

3 (bases 1 to 185722)
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collamore,A., Cook,A., Cooke,P., Corum,B., DeArillano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Strange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (18-JUN-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 18 2003 this sequence version replaced gi:22380811.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RN/RepeatMasker.html

TITLE
JOURNAL

COMMENT

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submission@genome.wi.mit.edu
 ----- Project Information
 Center project name: L18997
 Center clone name: 159.D.6
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 183868 bases at least Q40
 Consensus quality: 184372 bases at least Q30
 Consensus quality: 184505 bases at least Q20
 Insert size: 178000; agarose-ff
 Insert size: 185022; sum-of-contigs
 Quality coverage: 9.6 in Q20 bases; agarose-ff
 Quality coverage: 9.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 704: contig of 704 bp in length
 * 705
 805 804: gap of 100 bp
 * 805
 1472 1471: contig of 667 bp in length
 * 1472
 1572 1571: gap of 100 bp
 * 1572
 3159 3158: contig of 1587 bp in length
 * 3159
 3259 3258: gap of 100 bp
 * 3259
 11073 11072: contig of 7814 bp in length
 * 11073
 11173 29850: contig of 18678 bp in length
 * 11173
 29851 29950: gap of 100 bp
 * 29851
 70132 70131: contig of 40181 bp in length
 * 70132
 70232 70231: gap of 100 bp
 * 70232
 131332 131331: contig of 61100 bp in length
 * 131332
 131432 185722: contig of 54291 bp in length.
 * 131432
 Location/Qualifiers
 1. .185722

FEATURES
source

/organism="Mus musculus"
 /mol_type="Genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-159D6"
 /clone_lib="RPCI-23 Female Mouse BAC"
 1. .704
 /note="assembly_fragment"
 805. .1471
 /note="assembly_fragment"
 1572. .3158
 /note="assembly_fragment"
 3259. .11072
 /note="assembly_fragment"
 11173. .29850
 /note="assembly_fragment"
 29951. .70131
 /note="assembly_fragment"
 70232. .131331
 /note="assembly_fragment"
 131432. .185722
 /note="assembly_fragment"
 clone_end:17
 vector_side:right"

ORIGIN

Query Match 5.8%; Score 27; DB 2; Length 185722;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACCTGCTGCTGCTGCTGCTGCTGCTGCT 77
 |||||
 Db 153507 AACGCTGCTGCTGCTGCTGCTGCTGCT 153481

RESULT 35

AC098642/c

LOCUS

DEFINITION

AC098642

sequence.

AC098642

VERSION

AC098642.5

KEYWORDS

HTG.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jul 25, 2002 this sequence version replaced gi:18201769.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the


```

assembly was confirmed by restriction digest.
FEATURES
  Location/Qualifiers
    1..188834
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /clone="RP23-27007"
      /clone_lib="RPC1-23"
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 0.0012; Length 188834;
  Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 79536 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 79510
RESULT 36
AC114715 190402 bp DNA linear HTG 15-NOV-2002
LOCUS Rattus norvegicus clone CH230-339N2, WORKING DRAFT SEQUENCE, 4
DEFINITION AC114715
ACCESSION AC114715
VERSION AC114715.5 GI:25008972
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE
  1 (bases 1 to 190402)
    Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
    Allen, C., Allen, H., Alibrooks, S., Amin, A., Anguiano, D.,
    Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
    Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
    Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
    Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
    Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
    Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J.,
    Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
    Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
    Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
    Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
    Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
    Fernandez, S., Finley, M., Flaggs, N., Forbes, L., Foster, M., Foster, P.,
    Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
    Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
    Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
    Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
    Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
    Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
    Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
    Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
    Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
    Lorensuewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J.,
    Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
    Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
    Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
    Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
    Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
    Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
    Nwaokelimeh, O., Okwuonu, G., Olarnngyagoon, A., Pal, S., Parks, K.,
    Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
    Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
    Puazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R.,
    Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
    Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
    Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
    Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
    Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, K., Sosa, J.,

```

```

Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, D.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlaczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 190402)
Worley, K.C.
Direct Submission
Submitted (11-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 190402)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23681968.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTHG
Center Clone name: CH230-339N2
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 173476 bases at least Q40
Consensus quality: 175237 bases at least Q30
Consensus quality: 176250 bases at least Q20
Estimated insert size: 172780; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 4 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
  be preserved.
* 1 172097: contig of 172097 bp in length
* 172098 172197: gap of unknown length
* 172198 187288: contig of 15091 bp in length
* 187289 187388: gap of unknown length
* 187389 188702: contig of 1314 bp in length
* 188703 188802: gap of unknown length
* 188803 190402: contig of 1600 bp in length.
  Location/Qualifiers
    1..190402
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
FEATURES
  source

```

```

/db xref="taxon:10116"
/clone="CH230-339N2"
915 1432
/notes="clone_boundary
clone_end:sp6
site:
end sequence:BZ165181"
133344 .133527
/notes="clone_boundary
clone_end:T7
site:
end sequence:BZ165180"
172198 .173628
/notes="wgs end_extension
clone_end:T7"

ORIGIN
Query Match 5.8%; Score 27; DB 2; Length 190402;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAA 80
|||||
Db 5116 GCTGCTGCTGCTGCTGCTGCTGCTTAA 51142

RESULT 37
AC132877
LOCUS
DEFINITION
AC132877 Mus musculus clone RP24-216J21, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION
VERSION AC132877.3 GI:28927732
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192539)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-216J21
Unpublished
2 (bases 1 to 192539)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,J., Mihova,T., Mlanga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192539)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,K., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

```

```

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,J., Mihova,T., Mlanga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Willson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2003 this sequence version replaced gi:28412043.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27122
Center clone name: 216_J21
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191304 bases at least Q40
Consensus quality: 191385 bases at least Q30
Consensus quality: 191502 bases at least Q20
Insert size: 191939; sum-of-contigs
Quality coverage: 9.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7935: contig of 7935 bp in length
* 7936 8035: gap of 100 bp
* 8036 10717: contig of 2682 bp in length
* 10718 10817: gap of 100 bp
* 10818 21133: contig of 10316 bp in length
* 21134 21233: gap of 100 bp
* 21234 37038: contig of 15805 bp in length
* 37039 37138: gap of 100 bp
* 37139 73158: contig of 36020 bp in length
* 73159 73258: gap of 100 bp
* 73259 134748: contig of 61490 bp in length
* 134749 134848: gap of 100 bp
* 134849 192539: contig of 57691 bp in length.
*
* Location/Qualifiers
* 1..192539
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP24-216J21"
* /clone_lib="RPCI-24 Male Mouse BAC"
* 1..7935
* /note="assembly_fragment
* clone end:SP6
* vector side:left"
* 8035..10717
* /note="assembly_fragment"
* 10818..21133
* /note="assembly_fragment"
*
FEATURES
source
misc_feature
misc_feature
misc_feature

```

```

misc_feature      21234..37038
                  /note="assembly_fragment"
misc_feature      37139..73158
                  /note="assembly_fragment"
misc_feature      73259..134748
                  /note="assembly_fragment"
misc_feature      134849..192539
                  /note="assembly_fragment"
                  clone_end:17
                  vector_side:right"

ORIGIN
Query Match      5.8%; Score 27; DB 2; Length 192539;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCT 77
      |||||
Db 93242 AACGCTGCTGCTGCTGCTGCTGCT 93268

RESULT 38
AC118589
LOCUS      AC118589      192629 bp      DNA      linear      HTG 11-MAR-2003
DEFINITION Mus musculus clone RP23-198G19, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
VERSION      AC118589
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 192629)
REFERENCE      Birren,B., Nusbaum,C. and Lander,E.
            Mus musculus, clone RP23-198G19
            Unpublished
            2 (bases 1 to 192629)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
            Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kellis,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (19-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 192629)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collamore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,

```

```

Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 11, 2003 this sequence version replaced gi:24850527.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 117611
Center clone name: 198 G.19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191046 Bases at least Q40
Consensus quality: 191531 Bases at least Q30
Consensus quality: 191698 bases at least Q20
Insert size: 225000; agarose-fp
Quality coverage: 8.3 in Q20 bases; sum-of-contigs
Quality coverage: 9.7 in Q20 bases; sum-of-contigs
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 117690: contig of 117690 bp in length
* 117691 117790: gap of 100 bp
* 117791 118991: contig of 1201 bp in length
* 118992 119091: gap of 100 bp
* 119092 129620: contig of 10529 bp in length
* 129621 129720: gap of 100 bp
* 129721 147128: contig of 17407 bp in length
* 147128 147227: gap of 100 bp
* 147228 186320: contig of 39092 bp in length
* 186320 186419: gap of 100 bp
* 186420 192629: contig of 6210 bp in length.
FEATURES
            Location/Qualifiers
            1..192629
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone="RP23-198G19"
            /clone_lib="RPCT-23 Female Mouse BAC"
            1..117690
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left"
            117791..118991
            /note="assembly_fragment"
            119092..129620
            /note="assembly_fragment"
            129621..147128
            /note="assembly_fragment"
            147129..147227
            /note="assembly_fragment"
            186420..192629
            /note="assembly_fragment"

```

```

/misc_feature /note="assembly fragment"
147228..186319
/note="assembly fragment"
/misc_feature 186420..192629
/note="assembly fragment"
clone_end:T7
vector_side:right"

ORIGIN
Query Match 5.8%; Score 27; DB 2; Length 192629;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGTGTCTGCTGCTGCTGCTGCTTA 79
|||||
Db 144931 CGTGTCTGCTGCTGCTGCTGCTTA 144957

RESULT 39
AL929249/c
LOCUS AL929249 194209 bp DNA linear ROD 21-MAY-2003
DEFINITION Mouse DNA sequence from clone RP23-273G23 on chromosome 2, complete
sequence.
ACCESSION AL929249
VERSION AL929249.15 GI:31043747
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194209)
Direct Submission
Submitted (21-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 22, 2003 this sequence version replaced gi:30842758.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-273G23 is
from the RPCI-23 Mouse BAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1..194209
/organism="Mus musculus"

FEATURES
source

```

```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-273G23"
/clone_lib="RPCI-23"

Query Match 5.8%; Score 27; DB 10; Length 194209;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 AACGCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 118602 AACGCTGCTGCTGCTGCTGCTGCT 118576

RESULT 40
AL845432/c
LOCUS AL845432 195835 bp DNA linear ROD 02-DEC-2002
DEFINITION Mouse DNA sequence from clone RP23-332H19 on chromosome 2, complete
sequence.
ACCESSION AL845432
VERSION AL845432.7 GI:25955789
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 195835)
Direct Submission
Submitted (30-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 29, 2002 this sequence version replaced gi:25137036.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-332H19 is
from the RPCI-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1..195835
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

FEATURES
source

```

```

/chrnosome="2"
/clone="RP23-332H19"
/clone_lib="RPCI-23"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 195835;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 38844 AACGCTGCTGCTGCTGCTGCTGCTGCT 38818

RESULT 41
AC124555
LOCUS          210174 bp DNA linear ROD 08-NOV-2003
DEFINITION Mus musculus chromosome 14 clone RP23-245K17, complete sequence.
ACCESSION AC124555
VERSION AC124555.4 GI:38229416
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 210174)
The sequence of Mus musculus clone
Unpublished
REFERENCE 2 (bases 1 to 210174)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 210174)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 210174)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 8, 2003 this sequence version replaced gi:31880269.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M BA0245K17
----- Location/Qualifiers -----
1. .210174
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/clone="RP23-245K17"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 210174;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTAA 80
Db 206196 GCTGCTGCTGCTGCTGCTGCTGCTGCTAA 206222

RESULT 42
/chrnosome="2"
/clone="RP23-332H19"
/clone_lib="RPCI-23"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 195835;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 38844 AACGCTGCTGCTGCTGCTGCTGCTGCT 38818

RESULT 41
AC124555
LOCUS          210174 bp DNA linear ROD 08-NOV-2003
DEFINITION Mus musculus chromosome 14 clone RP23-245K17, complete sequence.
ACCESSION AC124555
VERSION AC124555.4 GI:38229416
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 210174)
The sequence of Mus musculus clone
Unpublished
REFERENCE 2 (bases 1 to 210174)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 210174)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 210174)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 8, 2003 this sequence version replaced gi:31880269.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M BA0245K17
----- Location/Qualifiers -----
1. .210174
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/clone="RP23-245K17"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 210174;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTAA 80
Db 206196 GCTGCTGCTGCTGCTGCTGCTGCTGCTAA 206222

RESULT 42
AL603682
LOCUS          211680 bp DNA linear ROD 24-OCT-2002
DEFINITION Mouse DNA sequence from clone RP23-235J5 on chromosome 11, complete
sequence.
ACCESSION AL603682
VERSION AL603682.36 GI:24394858
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211680)
Hopkins,B.
Direct Submission
Submitted (23-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquary@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:21952987.
----- Genome Center -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquary@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VBCOR: pBACe3.6.
----- Location/Qualifiers -----
1. .211680
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-235J5"
/clone_lib="RPCI-23"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 211680;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 25943 AACGCTGCTGCTGCTGCTGCTGCTGCT 25969

RESULT 43
AL604024
LOCUS          217278 bp DNA linear ROD 22-MAY-2002
DEFINITION Mouse DNA sequence from clone RP23-421F23 on chromosome 11,
complete sequence.
ACCESSION AL604024
VERSION AL604024.7 GI:21211848
KEYWORDS HTG.

```

```

SOURCE
ORGANISM      Mus musculus (house mouse)
REFERENCE
AUTHORS       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE         Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL       1 (bases 1 to 217278)
              Whitehead,S.
              Direct Submission
              Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              humquere@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
              On May 25, 2002 this sequence version replaced gi:18477353.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence was finished as follows unless otherwise noted: all
              regions were either double-stranded or sequenced with an alternate
              chemistry or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by at least
              one plasmid subclone or more than one M13 subclone; and the
              assembly was confirmed by restriction digest. The following
              abbreviations are used to associate primary accession numbers given
              in the feature table with their source databases: Em:, EMBL; Sw:,
              SWISSPROT; Tr:, TREMBL; Mp:, WORMPEP; Information on the WORMPEP
              database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
              from the RPL-23 Mouse PAC Library
              constructed by the group of Pieter de Jong.
              For further details see http://www.chori.org/bacpac/home.htm
              VECTOR: pBACe3.6.
FEATURES
source
1..217278
   /organism="Mus musculus"
   /db_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="11"
   /clone="RP23-421F23"
   /clone_lib="RPCI-23"
ORIGIN
Query Match      5.8%; Score 27; DB 10; Length 217278;
Best local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AACGCTGCTGCTGCTGCTGCTGCT 77
Db 58931 AACGCTGCTGCTGCTGCTGCTGCT 58957

RESULT 44
AC116661/c
LOCUS            237094 bp DNA linear HTG 13-SEP-2002
DEFINITION      Mus musculus chromosome UNK clone RP23-11601, WORKING DRAFT
SEQUENCE        16 unordered pieces.
AC116661
VERSION         AC116661.3 GI:22830524
KEYWORDS        HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
REFERENCE
AUTHORS         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL         1 (bases 1 to 237094)
AUTHORS         McPherson,J.D. and Waterston,R.H.
TITLE           The sequence of Mus musculus clone
JOURNAL         Unpublished
REFERENCE       2 (bases 1 to 237094)
AUTHORS         McPherson,J.D. and Waterston,R.H.
TITLE           Direct Submission
JOURNAL         Submitted (31-MAR-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA

```

```

3 (bases 1 to 237094)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (13-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Sep 13, 2002 this sequence version replaced gi:21105051.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0116001
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 230599 bases at least Q40
Consensus quality: 231829 bases at least Q30
Consensus quality: 232818 bases at least Q20
Insert size: 213000; agarose-fp
Quality coverage: 10.72 in Q20 bases; agarose-fp
Quality coverage: 9.73 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1799: contig of 1799 bp in length
* 1800: gap of unknown length
* 1900: contig of 1492 bp in length
* 3392: gap of unknown length
* 3491: contig of 1243 bp in length
* 4735: gap of unknown length
* 4834: contig of 1650 bp in length
* 6484: gap of unknown length
* 6585: contig of 3672 bp in length
* 10257: gap of unknown length
* 10357: contig of 6909 bp in length
* 17266: gap of unknown length
* 17365: contig of 7048 bp in length
* 24414: gap of unknown length
* 24513: contig of 8359 bp in length
* 32872: gap of unknown length
* 32873: contig of 8883 bp in length
* 41855: gap of unknown length
* 41956: contig of 13438 bp in length
* 55393: gap of unknown length
* 55494: contig of 19878 bp in length
* 75371: gap of unknown length
* 75372: contig of 31206 bp in length
* 75472: gap of unknown length
* 106677: contig of 30610 bp in length
* 106778: gap of unknown length
* 137387: contig of 46158 bp in length
* 137488: contig of 46158 bp in length
* 183646: gap of unknown length
* 183745: gap of unknown length
* 236859: contig of 53114 bp in length
* 236860: gap of unknown length
* 236960: contig of 135 bp in length.
*
* Location/Qualifiers
* 1..237094
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"

```

```

/misc_feature /chromosome="UNK"
/clonetype="FP23-11601"
1. .1799
/note="assembly_name:Contig46"
1900. .3391
/note="assembly_name:Contig47"
3492. .4734
/note="assembly_name:Contig48"
4835. .6484
/note="assembly_name:Contig49"
6585. .10256
/note="assembly_name:Contig50
clone end:77
vector side:right"
/misc_feature 10357. .17265
/note="assembly_name:Contig51"
17366. .24413
/note="assembly_name:Contig52"
24514. .32872
/note="assembly_name:Contig53"
32973. .41855
/note="assembly_name:Contig54"
41956. .55393
/note="assembly_name:Contig55"
55494. .75371
/note="assembly_name:Contig56"
75472. .106577
/note="assembly_name:Contig57"
106778. .137387
/note="assembly_name:Contig58"
137488. .183645
/note="assembly_name:Contig59"
183746. .236859
/note="assembly_name:Contig60"
236960. .237094
/note="assembly_name:Contig10"

Query Match 5.8%; Score 27; DB 2; Length 237094;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 AACCTGCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 165807 AACCTGCTGCTGCTGCTGCTGCT 165581

RESULT 45
AC131846/c AC131846 237332 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-227F3, WORKING DRAFT SEQUENCE, 3
DEFINITION
unorderd pieces.
ACCESSION AC131846
VERSION AC131846.4 GI:30579738
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 237332)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,J., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Georgiadis,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangun,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankweli,O., Okwunodu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,D.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Holt,R.A., O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 237332)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (26-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237332)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23606053.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with 'N's to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GLIQ
Center clone name: CH230-227F3
----- Summary Statistics

```


Assembly program: Atlas 3.0;
 Consensus quality: 203099 bases at least Q40
 Consensus quality: 207416 bases at least Q30
 Consensus quality: 210436 bases at least Q20
 Estimated insert size: 213682; sum-of-coverage estimation
 Quality coverage: 7x in Q20 bases; sum-of-coverage estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 63260: contig of 63260 bp in length
 * 63261 63360: gap of unknown length
 * 63361 235896: contig of 172536 bp in length
 * 235897 235996: gap of unknown length
 * 235997 237332: contig of 1336 bp in length.

FEATURES

Source

1..237332
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-227F3"

misc_feature

654..1459
 /note="clone boundary
 clone_end:Sp6
 site:EcORI"

misc_feature

8193..8812
 /note="clone boundary
 clone_end:T7
 site:EcORI"

misc_feature

63361..65231
 /note="wgs end_extension
 clone_end:T7"

ORIGIN

Query Match 5.8%; Score 27; DB 2; Length 237332;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTTAA 80
 |||||
 Db 5161 GCTGCTGCTGCTGCTGCTGCTTAA 5135
 |||||

RESULT 46

AC099207/c
 LOCUS Rattus norvegicus clone CH230-132D18, WORKING DRAFT SEQUENCE, 2
 DEFINITION Rattus norvegicus clone CH230-132D18, WORKING DRAFT SEQUENCE, 2
 UNORDERED PIECES.

AC099207

AC099207.6 GI:30521989
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 240501)
 Muzny, D., Marie, Metzker, M., Lee, A., Brannan, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Altschuler, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Dwyer-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C.,
 Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshewala, L., Louis, H., Lozano, R., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M., McNeill, T., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackeleme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
 Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D., Holt, R., Smith, H., O.,
 Weinstock, G., and Gibbs, R.A.

Direct Submission

Unpublished
 2 (bases 1 to 240501)

Worley, K.C.

Direct Submission

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240501)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:25095914.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center project name: GHS

Center clone name: CH230-132D18

----- Summary Statistics -----

Assembly program: Atlas 3.0;

Consensus quality: 227593 bases at least Q40

Consensus quality: 229405 bases at least Q30

Consensus quality: 230440 bases at least Q20

Estimated insert size: 236821; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

* This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 238878: contig of 238878 bp in length

* 238879 238978: gap of unknown length

* 238979 240501: contig of 1523 bp in length.

FEATURES

source

1. 240501

Location/Qualifiers

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clones="CH230-132D18"

1. 1103

misc_feature

/note="wgs_contig"

ORIGIN

Query Match 5.8%; Score 27; DB 2; Length 240501;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCT 77

|||||

DB 47005 AACGCTGCTGCTGCTGCTGCTGCT 46979

RESULT 47

AC129456/c

LOCUS

AC129456 Rattus norvegicus clone CH230-9K8, WORKING DRAFT SEQUENCE.

DEFINITION

AC129456

ACCESSION

AC129456.5 GI:24940997

VERSION

HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 240561)

Muzzy, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hawes, A., Hamilton, C., Hamilton, C., Hamilton, K., Harvey, I., Havlak, P., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., King, Z., King, Z., King, Z., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenschwar, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, P., Mapua, P., Martin, K., Martin, R., Martinez, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L., Plopper, F., Poindexter, A., Popovic, D., Reeves, K., Regier, M. A., Reigh, R., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Richards, S., Riggs, F., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Taylor, C., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 240561)

Worley, K. C.

Direct Submission

Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240561)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:23096544.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center project name: GDS

Center clone name: CH230-9K8

----- Summary Statistics -----

Assembly program: Phrap; version 0.990329

Consensus quality: 212559 bases at least Q40

Consensus quality: 214527 bases at least Q30

Consensus quality: 215680 bases at least Q20

[illegible]

```

Query Match          5.8%; Score 27; DB 2; Length 250546;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTTAA 80
Db 14787 GCTGCTGCTGCTGCTGCTGCTTAA 14813

RESULT 50
AC128474
LOCUS
DEFINITION
AC128474
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC128474 258061 bp DNA linear HTG 21-SEP-2002
Rattus norvegicus clone CH230-51A3, *** SEQUENCING IN PROGRESS ***.
AC128474.2 GI:23265072
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 258061)
Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anylebechi,V., Ayogi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,S., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gumaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,J., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hui,J., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Louisegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankweme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poinexter,A., Popovic,D., Primus,E., Pu.L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajis,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhen,D., Zoon,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 258061)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 258061)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 21, 2002 this sequence version replaced gi:21909191.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
-----
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GYUH
Center clone name: CH230-51A3
-----
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 237914 bases at least Q40
Consensus quality: 240542 bases at least Q30
Consensus quality: 242364 bases at least Q20
Estimated insert size: 259476; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 258061: contig of 258061 bp in length.
FEATURES
source
Location/Qualifiers
1..258061
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-51A3"
1..1429
/note="wgs end extension
clone_end:T7"
8852..9636
/note="clone boundary
clone_end:T7"
site:ECORI
end_sequence:BH336930"
248189..248838
/note="clone boundary
clone_end:Sp6"
site:ECORI
end_sequence:BH336932"
256785..258061
/note="wgs end extension
clone_end:Sp6"
misc_feature
misc_feature
misc_feature
misc_feature
ORIGIN

```

Query Match 5.8%; Score 27; DB 2; Length 258061;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 51 AACGCTGCTGCTGCTGCTGCTGCT 77
Db 82726 AACGCTGCTGCTGCTGCTGCTGCT 82752

Search completed: May 25, 2004, 18:30:57
Job time : 3357 secs